1

A Bayesian Technique for Image Classifying Registration

Mohamed Hachama*, Agnès Desolneux and Frédéric J.P. Richard

Abstract

In this paper, we address a complex image registration issue arising when the dependencies between intensities of images to be registered are not spatially homogeneous. Such a situation is frequently encountered in medical imaging when a pathology present in one of the images modifies locally intensity dependencies observed on normal tissues. Usual image registration models, which are based on a single global intensity similarity criterion, fail to register such images, as they are blind to local deviations of intensity dependencies. Such a limitation is also encountered in contrast enhanced images where there exist multiple pixel classes having different properties of contrast agent absorption. In this paper, we propose a new model in which the similarity criterion is adapted locally to images by classification of image intensity dependencies. Defined in a Bayesian framework, the similarity criterion is a mixture of probability distributions describing dependencies on two classes. The model also includes a class map which locates pixels of the two classes and weights the two mixture components. The registration problem is formulated both as an energy minimization problem and as a Maximum A Posteriori (MAP) estimation problem. It is solved using a gradient descent algorithm. In the problem formulation and resolution, the image deformation and the class map are estimated at the same time, leading to an original combination of registration and classification that we call image classifying registration. Whenever sufficient information about class location is available in applications, the registration can also be performed on its own by fixing a given class map. Finally, we illustrate the interest of our model on two real applications from medical

M. Hachama is with the University of Khemis Miliana, Laboratory of Energy and Intelligent Systems, Route Theniat El Had, Algeria, e-mail:hachamam@gmail.com

A. Desolneux is with the ENS Cachan, CMLA, CNRS UMR 8536, 61, avenue du président Wilson, 94235 Cachan. e-mail: desolneux@cmla.ens-cachan.fr

F. Richard is with the University of Provence, LATP, CNRS UMR 6632, 39, rue F. Joliot Curie, 13453 Marseille Cedex 13, France, e-mail: richard@cmi.univ-mrs.fr

imaging: template-based segmentation of contrast-enhanced images and lesion detection in mammograms. We also conduct an evaluation of our model on simulated medical data and show its ability to take into account spatial variations of intensity dependencies while keeping a good registration accuracy.

Index Terms

Image registration, mixture models, lesion detection.

I. Introduction

Image registration is a central issue of image processing, which is particularly encountered in medical applications [1]–[5]. Medical image registration is critical for the fusion of complementary information about patient anatomy and physiology, for the longitudinal study of a human organ over time and the monitoring of disease development or treatment effect, for the statistical analysis of a population variation in comparison to a so-called digital atlas, for image-guided therapy, etc.

Image registration consists in mapping domains of several images onto a common space and results in some corrections of geometric differences between the images. Most of classical registration techniques rely upon the assumption that there exists a relationship between intensities of images to be registered and that this relationship remains the same all over the image domains [6]–[11]. This assumption is typically made when applying registration techniques based on intensity criteria such as the sum of squared differences, the correlation ratio, the correlation coefficient or the mutual information [10].

But such an assumption is not always satisfied. As an example, let us mention the medical imaging case when a contrast agent is used to enhance some pathological tissues (lesions) [12], [13]. After enhancement, intensities of normal tissues and lesions are likely to differ, even though they can be the same before enhancement. So, a same intensity before enhancement may correspond to several intensities after enhancement. Hence, with contrast-enhanced imaging modalities, the relationship between image intensities is neither unique, nor spatially invariant. It mainly depends on the type of observed tissues. In such cases, ignoring the spatial context may lead to locally establishing an inaccurate or even inconsistent correspondence between homologous geometric structures. This issue was documented in [13], [14], where it is shown that such non-rigid registration would wrongly change the size of non-deformed contrast-enhanced structures.

In the literature, there have been several works dealing with image registration in the presence of multiple pixel classes. These works can mainly be divided into two categories: those based on robust estimation and mixture models, and those combining registration and classification (or segmentation).

Robust estimation is a statistical approach which has been widely applied to image processing [15]. This approach involves the definition of outliers, which are characterized as elements deviating from a normal model, detected and possibly rejected [16]–[19]. Applied to optical flow estimation and image registration [20]–[25], robust estimation helps to reduce the influence of large insignificant image differences on the optical flow or the deformation estimation. However, these approaches offer poor characterizations of outliers, which are usually described as pixels generating large image differences. They cannot deal with complex situations arising from medical imaging applications.

More general robust estimation approaches are based on mixture models [18] and characterize outliers by some specific probability distributions [26]–[34]. In optical flow computation, a mixture model was used to distinguish between several layers of movements and an outlier class [31]. In image registration, some authors used a mixture model in which image differences generated by outliers are represented by a mixture component [29], [30]. Similar approaches could be used in medical image registration considering medical lesions as outliers [26]–[28]. However, the main and important difference with the model we introduce in this paper is that the mixture models proposed above do not use any spatial and geometric information about the pixel classes but only pixel-independent mixing proportions.

In other approaches, spatial informations derived from segmentation were used to adapt regionally the similarity criterion describing intensity relationships [35]–[38]. Such segmentation-based approaches require a preliminary segmentation of the images which, obviously, cannot always be obtained. For instance, in dynamic contrast-enhanced sequences, the image segmentation has to be estimated from the registered images [39].

Image segmentation has also been combined with atlas-based registration in Bayesian frameworks [40], [41], where mixture models were used to represent pixel intensities of different anatomical structures. In other works, Markov random fields were employed to describe a label map of pixel classes. The registration and the segmentation were then computed by the Maximum A Posteriori (MAP) estimation [42]–[44]. In such approaches, one can incorporate prior information about the lesion shape and localization. However, the proposed methods used simple characterizations of intensity variations.

In this paper, we deal with the issue of registering images whose intensity relationship is spatially dependent. We propose a new technique where the registration is simultaneously combined to a pixel classification. This classification provides some spatial information about intensity relationships. We use mixture models to take into account complex intensity changes and Markov random fields to label pixels.

The paper is organized as follows. Section 2 describes the theoretical foundation of the proposed new method. Then, the numerical aspects are discussed in Section 3. In Section 4, experiments are conducted

both on simulated and on real data to evaluate the method performance. Finally, some conclusions and possible extensions of the proposed approach are discussed in the last section.

II. The proposed method

A. The Bayesian framework

Let m and n be two integers, and $\Omega_d = \{0, ..., m-1\} \times \{0, ..., n-1\}$ be a discrete grid of size $N = m\,n$. Observed images are real-valued functions defined on Ω_d. They can be interpolated on the continuous domain $\Omega = [0, m-1] \times [0, n-1]$ naturally associated to the grid Ω_d. For convenience, elements of Ω_d are ordered and will be denoted x_i for $i = 1, \cdots, N$.

The registration of two images I and J consists in finding a mapping $\phi : \Omega \to \Omega$ for which the deformed image $I^\phi := I \circ \phi$ of the so-called source image I is as similar as possible to the so-called target image J.

In a Bayesian framework, the mapping ϕ is usually obtained as a MAP estimate (see [45] and references therein). Specific to this framework, images and deformations are assumed to be some realizations of random fields indexed on Ω_d and Ω, respectively. For each point x_i of Ω_d, $I(x_i)$ and $J(x_i)$ are realizations of two real-valued random variables and, for each point x of Ω, $\phi(x)$ is a realization of a random vector with values in Ω. The relationships between the intensities of the registered images are then statistically described by a probability distribution of J given I, ϕ and a set of parameters θ (this conditional distribution is denoted by $\pi(J \mid I, \phi; \theta)$). Usually, the variables $J(x_i)$ are assumed to be independent conditionally to the variables $I^\phi(x_i)$. Hence, $\pi(J \mid I, \phi; \theta)$ can be written as

$$\pi(J \mid I, \phi; \theta) = \prod_{i=1}^{N} \pi(J(x_i) \mid I^\phi(x_i); \theta). \tag{1}$$

Because of noise, and also because it is a very generic choice, it is possible to assume that the intensity differences between $J(x)$ and $I^\phi(x)$ follow a Gaussian distribution with mean μ and variance σ^2 at each pixel $x \in \Omega_d$, leading to the distribution

$$\pi(J(x) \mid I^\phi(x); \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(J(x) - I^\phi(x) - \mu)^2\right), \tag{2}$$

with $\theta = (\mu, \sigma)$. In this definition, it is worth noticing that the mean μ and the variance σ^2 do not depend on any position x. Consequently, the intensity relationship between images is spatially homogeneous. Let us further mention that, similarly to this particular Gaussian distribution, distributions associated to other usual criteria (e.g. correlation ratio, correlation coefficient, or mutual information) are also homogeneous.

Besides, a prior probability distribution is set on the mappings ϕ in order to favour smooth deformations as solutions of the registration problem. Let $\phi = \mathrm{id} + u$ be written as the sum of the identity map id and a displacement field u, seen as an element of a Hilbert space $\mathcal{H}$ equipped with an inner product $a(\cdot,\cdot)$. Let then $\tilde{\mathcal{H}}$ be a finite-dimensional approximation subspace of $\mathcal{H}$ which is spanned by a basis $\mathcal{B} = (\psi_i)_{i=1}^{n_e}$, with $n_e \in \mathbb{N}^*$. In $\tilde{\mathcal{H}}$, a displacement u is uniquely determined by a vector of coefficients $\mathbf{b} = (\beta_i)_{i=1}^{n_e}$ such that $u = \sum_{i=1}^{n_e} \beta_i \cdot \psi_i$. In the following, the deformations ϕ are assumed to be expanded into the subspace $\tilde{\mathcal{H}}$ and identified with their decomposition coefficients $\mathbf{b}$. We will also use the notation $I^{\mathbf{b}}$ instead of I^{ϕ}. Several choices are possible for the basis $\mathcal{B}$: Fourier basis, sine and cosine transform basis functions, B-splines, piecewise affine or trilinear basis functions, wavelets (see [46] for a review), etc.

We consider an inner product on $\tilde{\mathcal{H}}$, given by $a(u,u) = \mathbf{b}^T A \mathbf{b}$, where $\mathbf{b}^T$ is the transpose of $\mathbf{b}$, and A is a symmetric positive-definite matrix. On $\tilde{\mathcal{H}}$, we then define a centered multivariate normal distribution with covariance matrix A^{-1} given by

$$\pi(\mathbf{b}) = (2\pi)^{\frac{-n_e}{2}} \sqrt{\det(A^{-1})} e^{-\frac{1}{2}\mathbf{b}^T A \mathbf{b}}. \tag{3}$$

This prior distribution is used as a regularity term to enforce the smoothness of the deformations ϕ.

A usual choice for $a(\cdot,\cdot)$ is the bilinear form

$$a(u,v) = \frac{1}{2}\int_\Omega \sum_{i,j=1}^{2} \frac{\lambda}{2}\left(\frac{\partial u_i(x)}{\partial y_i}\frac{\partial v_j(x)}{\partial y_j}\right) dy + \frac{1}{2}\int_\Omega \sum_{i,j=1}^{2} \frac{\mu}{4}\left(\frac{\partial u_i(x)}{\partial y_j} + \frac{\partial v_j(x)}{\partial y_i}\right)^2 dy, \tag{4}$$

defined for $\lambda > 0$ and $\mu > 0$, which is an inner product on the Sobolev space $H^1(\Omega;\mathbb{R}^2)$ and is related to the linearized strain energy of elastic materials [47]–[50]. We choose here the elastic regularization without aiming at a specific application, but knowing that this term enables to deal with many medical imaging applications. Other regularity terms could be better adapted to specific applications. They can be easily introduced in the proposed model. Other choices include the membrane energy [51], the bending energy [52], etc. The matrix A can also be estimated from the data using for instance EM algorithms with stochastic approximation [53], [54].

In the model we propose, we will use a basis $\mathcal{B}$ constructed using the finite element method (the details of the construction will be given in Section III) and the inner product given by Equation (4).

The registration problem is defined as the problem of finding the MAP, *i.e.* maximizing the posterior distribution $\pi(\mathbf{b}|I,J;\theta)$ with respect to the coefficients $\mathbf{b}$. Using Bayes' Theorem, this problem can be

equivalently formulated as finding

$$\mathbf{b}^\star = \arg\max_{\mathbf{b}} \pi(J|I,\mathbf{b};\theta)\ \pi(\mathbf{b}). \tag{5}$$

Applying the $-\log$ function to all probability distributions, this maximization problem can also be transformed into the problem of minimizing the energy

$$E(\mathbf{b}) = S(I,J,\mathbf{b};\theta) + H(\mathbf{b}), \tag{6}$$

where $S(I,J,\mathbf{b};\theta) = -\log \pi(J|I,\mathbf{b};\theta)$ and $H(\mathbf{b}) = -\log \pi(\mathbf{b})$.

B. The two-class registration model

We now outline our new model. It is an extension of the model described in the previous section. The main feature is the introduction and the estimation of a pixel classification to take into account the spatial variations of the statistical relationships between the intensities of J and I^ϕ.

1) Intensity relationships: In Equation (2), the probability distribution $\pi(J(x)|I^{\mathbf{b}}(x);\theta)$ describing the intensity relationship is spatially homogeneous. We now assume that the pixels of image J can be divided into two classes (labeled 0 and 1) where the intensity relationships are different and denoted respectively by $\pi_j(J(x)|I^{\mathbf{b}}(x);\theta_j)\,,\ j \in \{0,1\}$. Let also $L(x)$ be the probability for a pixel x to belong to the class 1. Then, the intensity relationship at pixel x is described by the mixture distribution

$$\pi(J(x)|I^{\mathbf{b}}(x),L(x);\theta) = \pi_0(J(x)|I^{\mathbf{b}}(x);\theta_0)\ (1-L(x)) + \pi_1(J(x)|I^{\mathbf{b}}(x);\theta_1)\ L(x), \tag{7}$$

where $\theta = (\theta_0,\theta_1)$.

Let us denote $L_i = L(x_i)$ and $\mathbf{L} = (L_1,\cdots,L_N)^T$ the vector of class probabilities on grid pixels. Assuming the conditional independence on grid points, we obtain the global conditional distribution

$$\pi(J|I^{\mathbf{b}},\mathbf{L};\theta) = \prod_{i=1}^{N} \Big(\pi_0(J(x_i)|I^{\mathbf{b}}(x_i);\theta_0)\ (1-L_i) + \pi_1(J(x_i)|I^{\mathbf{b}}(x_i);\theta_1)\ L_i\Big). \tag{8}$$

In an application of our model to lesion detection, the classification aims at distinguishing pixels on a lesion (class 1) from those outside any lesion (class 0). In contrast-enhanced images, the enhancement is more important on lesions than it is on normal tissues, leading to higher image differences on lesions. Assuming that the distributions are Gaussian on both classes, we can define distributions π_0 and π_1 as

in Equation (2) using two different sets of parameters: $\theta_0 = (\mu_0, \sigma_0)$ for π_0 and $\theta_1 = (\mu_1, \sigma_1)$ for π_1, with $\mu_1 > \mu_0$. Another possibility is, as in robust estimation, to consider elements of class 1 as outliers of class 0 and put no specific information on class 1 by defining π_1 as the uniform distribution

$$\pi_1(J(x)|I^{\mathbf{b}}(x); \theta_1) = \frac{1}{N_{\text{gl}}}, \tag{9}$$

where N_{gl} is the number of possible gray-levels of J.

Finally, we define a prior distribution on the class map $\mathbf{L}$ itself in order to introduce some constraints on the spatial homogeneity of the pixels of a same class. For that purpose, we equip the grid Ω_d with a neighborhood system (either a 4 or 8 neighborhood system) and assume that the class map $\mathbf{L}$ is a Markov random field on it [55]. We have considered two prior distributions on L. The first one is a Gaussian model, given by

$$\pi(\mathbf{L}) = \frac{1}{Z} \exp\left(-\alpha_1 \sum_{i=1}^{N} L_i^2 - \alpha_2 \sum_{\{i,j=1,\cdots,N; x_i \sim x_j\}} (L_i - L_j)^2 \right), \tag{10}$$

where $x_i \sim x_j$ means that x_i and x_j are neighboring pixels, Z is a normalization constant, $\alpha_1 > 0$ and $\alpha_2 > 0$. The second model, which is a particular case of the Gaussian model when L is binary with range $\{0, 1\}$, is the Bernoulli model

$$\pi(\mathbf{L}) = \frac{1}{Z} \exp\left(-\alpha_1 \sum_{i=1}^{N} L_i + \alpha_2 \sum_{\{i,j=1,\cdots,N; x_i \sim x_j\}} L_i L_j \right), \tag{11}$$

where $\alpha_1 > 0$ and $\alpha_2 > 0$. If we let $X = 2L - 1$, this model is equivalent to the Ising model [56]. In models (10) and (11), the parameter α_1 restricts the amount of pixels of the class 1, whereas the parameter α_2 enforces the spatial homogeneity of the classes.

2) Classes and deformations: In some situations, one must locally adapt the deformation field to structures that must be kept rigid, using a tissue-dependent filtering technique. Otherwise, rigid tissue, such as bone, could be deformed elastically, growth of tumors may be concealed, and contrast-enhanced structures may be reduced in volume [57]. Such a specificity could be taken into account in our Bayesian framework by defining a distribution on deformations that would depend on classes: setting distributions $\pi_0(\mathbf{b})$ and $\pi_1(\mathbf{b})$ respectively for the two classes, we could define a distribution $\pi(\mathbf{b}; L)$ as a mixture of

the distributions $\pi_0(\mathbf{b})$ and $\pi_1(\mathbf{b})$.

However, such an option was not considered in our applications: deformation differences between lesions and normal tissues are not significant and do not justify increasing the complexity of the model. Hence, in this work, we rather make the assumption that the deformations and the classes are independent. Further assuming that they are both independent from images, we get

$$\pi(\mathbf{b}, \mathbf{L} \mid I) = \pi(\mathbf{b})\ \pi(\mathbf{L})$$

3) MAP estimations: In our Bayesian framework, the registration problem consists in maximizing the posterior distribution $\pi(\mathbf{b}, \mathbf{L}|I, J; \theta)$ with respect to the deformation $\mathbf{b}$ and to the class map $\mathbf{L}$. In this new formulation, registration and classification are both integrated in a single maximization problem, and have to be performed simultaneously.

Using Bayes' Theorem, it comes that

$$\begin{aligned} \pi(\mathbf{b}, \mathbf{L}|I, J; \theta) &= \frac{\pi(J|I, \mathbf{L}, \mathbf{b}; \theta)\ \pi(\mathbf{b}, \mathbf{L}|I)}{\pi(J|I; \theta)} \\ &= \frac{\pi(J|I, \mathbf{L}, \mathbf{b}; \theta)\ \pi(\mathbf{b})\ \pi(\mathbf{L})}{\pi(J|I; \theta)} \end{aligned} \tag{12}$$

Hence the registration problem can be restated in terms of likelihood and prior distributions as

$$(\widetilde{\mathbf{b}}, \widetilde{\mathbf{L}}) = \arg\max_{\mathbf{b},\mathbf{L}}\ \pi(J|I, \mathbf{L}, \mathbf{b}; \theta)\ \pi(\mathbf{b})\ \pi(\mathbf{L}). \tag{13}$$

Again, by application of the $-\log$ transform to distributions, this maximization problem can be transformed into the problem of minimizing the energy

$$\mathcal{E}(\mathbf{b}, \mathbf{L}) = S(I, J, \mathbf{b}, \mathbf{L}; \theta) + H(\mathbf{b}) + R(\mathbf{L}), \tag{14}$$

where $S(I, J, \mathbf{b}, \mathbf{L}; \theta) = -\log \pi(\mathbf{b}, \mathbf{L}|\theta, I, J)$, $R(\mathbf{L}) = -\log \pi(\mathbf{L})$ and, $H(\mathbf{b}) = -\log \pi(\mathbf{b})$ is as defined in Equation (6).

Besides, when the class map L is known ($L \in [0, 1]^N$), it is also possible to only solve the registration problem

$$\widetilde{\mathbf{b}} = \arg\max_{\mathbf{b}}\ \pi(J|I, \mathbf{L}, \mathbf{b}; \theta)\ \pi(\mathbf{b}). \tag{15}$$

This leads to a registration with two classes and can be used for template-based segmentation (see Section IV-B1).

4) Parameters: The distributions defined above involve several parameters. In our application, we used the prior deformation distribution defined by Equation (4) and set manually the Lame constants λ and μ from experiments. We also set manually the mean μ_0 and the variance σ_0 of the Gaussian distribution π_0 of the class 0, and the weights α_1 and α_2 of the Gaussian or Bernoulli model in Equations (10) and (11). The other parameters, which are the mean μ_1 and the variance σ_1 of the Gaussian distribution π_1 on class 1, are estimated from the data. We assume that μ_1 belongs to an interval $[\mu_{min}, \mu_{max}]$ and σ_1 to another interval $[\sigma_{min}, \sigma_{max}]$ and put, as a prior, uniform independent distributions of μ_1 and σ_1 on these intervals. The parameter vector θ is decomposed as $\theta = (\theta_0, \theta_1)$, where $\theta_0 = (\mu_0, \sigma_0)$ is known and $\theta_1 = (\mu_1, \sigma_1)$. This parameter θ_1 is then estimated, together with the deformation and the classification by solving the following MAP estimation problem:

$$(\widetilde{\mathbf{b}}, \widetilde{\mathbf{L}}, \widetilde{\theta}_1) = \arg\max_{\mathbf{b},\mathbf{L},\theta_1} \pi(J|I, \mathbf{L}, \mathbf{b}; \theta_0, \theta_1)\, \pi(\mathbf{b})\, \pi(\mathbf{L})\, \pi(\theta_1). \tag{16}$$

As before, this is equivalent to the minimization of an energy of the form

$$\mathcal{E}(\mathbf{b}, \mathbf{L}, \theta_1) = S(I, J, \mathbf{b}, \mathbf{L}; \theta) + H(\mathbf{b}) + R(\mathbf{L}), \tag{17}$$

under the constraint that $\mu_1 \in [\mu_{min}, \mu_{max}]$ and $\sigma_1 \in [\sigma_{min}, \sigma_{max}]$.

III. Numerical resolution

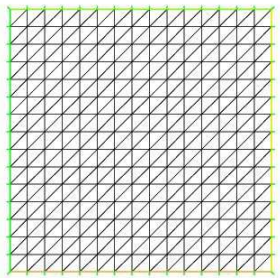
In this section, we present the numerical resolution of the combined registration and classification problem defined in Equation (16) or equivalently in Equation (17), using the gradient descent algorithm.

A. Problem discretization

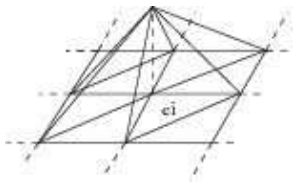
The discretization of the problem is based on the construction of a finite dimensional basis of deformations $\mathcal{B}$ using the finite element method [58].

First, the domain Ω is partitioned into a set of triangles whose vertices are in a set denoted $\Omega_h = \{e_i \in \Omega, 1 \leq i \leq n_e\}$. Such a partition is obtained by applying a Delaunay triangulation to the continuous domain Ω. Then, a $P1$-Lagrange finite element function Φ_i is associated to each vertex e_i where Φ_i is the only piecewise affine function on the triangulated domain satisfying $\Phi_i(e_j) = 1$ if $i = j$ and 0 otherwise. The function Φ_i is locally supported on triangles of the partition having e_i as vertex (Figure 1).

Next, we define $\psi_i^1 = (\Phi_i, 0)$, $\psi_i^2 = (0, \Phi_i)$, and the family of functions $\mathcal{B} = \{\psi_i^k, i = 1, ..., n_e, k = 1, 2\}$. The family $\mathcal{B}$ spans a finite dimensional space $\tilde{H}^1$ that is a subset of the infinite dimensional



(a)



(b)

Fig. 1. Discretization of the deformations domain Ω: (a) The mesh, (b) a $P1$-Lagrange finite element function Φ_i is associated to a vertex e_i (piecewise affine function).

Sobolev space $H^1(\Omega;\mathbb{R}^2)$. In $\tilde{H}^1$, a displacement u can be written as

$$u = \sum_{i=1}^{n_e} (\beta_i^1 \psi_i^1 + \beta_i^2 \psi_i^2) = \sum_{i=1}^{n_e} \beta_i \cdot \psi_i, \tag{18}$$

where $\psi_i = \psi_i^1 + \psi_i^2$ and $\beta_i = (\beta_i^1, \beta_i^2)^T \in \mathbb{R}^2$. We denote $\mathbf{b} = (\beta_1^1, \beta_1^2, \cdots, \beta_{n_e}^1, \beta_{n_e}^2)^T$. Then, up to a constant, the energy of the deformation given by Equations (3) and (4) becomes

$$\begin{aligned} H(\mathbf{b}) &= \frac{1}{2} a(u,u) \\ &= \frac{1}{2} \sum_{i=1}^{n_e} \beta_i^T A_{ii} \beta_i + \sum_{i,j=1, i \neq j}^{n_e} \beta_i^T A_{ij} \beta_j, \end{aligned} \tag{19}$$

where $A_{ij} = \begin{pmatrix} a_{ij}^{11} & a_{ij}^{12} \\ a_{ij}^{21} & a_{ij}^{22} \end{pmatrix}$ are 2×2-matrices defined by $a_{ij}^{km} = a(\psi_i^k, \psi_j^m)$ for $k, m = 1, 2$. Choosing the elasticity potential in Equation (4), it can be shown that each matrix A_{ij} is symmetric and that $A_{ij} = A_{ji}$. Besides, since the functions ψ_i^k are locally supported on the triangles whose vertices contain e_i, the matrices A_{ij} are null whenever the vertices e_i and e_j are not on a same triangle. Hence, $H(\mathbf{b})$ reduces to

$$H(\mathbf{b}) = \frac{1}{2} \sum_{i=1}^{n_e} \beta_i^T A_{ii} \beta_i + \sum_{\{i,j=1, e_i \sim e_j\}}^{n_e} \beta_j^T A_{ij} \beta_i. \tag{20}$$

where $e_i \sim e_j$ indicates that e_i and e_j are different vertices of a same triangle.

B. Gradient descent

The energy $\mathcal{E}$ in Equation (17) is non-linear and non-convex with respect to $\mathbf{b}$ and θ. A first approach to minimize this energy is to approximate the solutions of the Euler equations associated to $\mathcal{E}$ using a gradient descent algorithm [59].

Let X denote the vector formed by all the variables

$$\begin{aligned} X &= (\mathbf{b}, \mathbf{L}, \theta_1) \\ &= (\beta_1, ..., \beta_{n_e}, L_1, ..., L_N, \mu_1, \sigma_1) \end{aligned}$$

The gradient descent algorithm consists in defining an initialization $X^{(0)}$ of X and iterating the scheme ($t > 0$)

$$X^{(t+1)} = X^{(t)} - \rho^{(t)} \cdot \nabla \mathcal{E}(X^{(t)}). \tag{21}$$

until stabilization; $\rho^{(t)}$ is a vector of the same size as X and with positive components. In this equation, $\nabla \mathcal{E}(.)$ is the gradient of $\mathcal{E}$ with respect to X. To compute this gradient, we need the expression of the partial derivatives of the energy. We have

$$\begin{aligned} \frac{\partial \mathcal{E}}{\partial \beta_k} = A_{kk}\beta_k + & \sum_{\{i=1,\cdots,n_e; e_i \sim e_k\}} A_{ki}\beta_i \\ & - \sum_{i=1}^{N} \left(p_i^0 \frac{\left(J(x_i) - I^{\mathbf{b}}(x_i) - \mu_0\right)}{\sigma_0^2} + p_i^1 \frac{\left(J(x_i) - I^{\mathbf{b}}(x_i) - \mu_1\right)}{\sigma_1^2} \right) \\ & \qquad .\Phi_k(x_i)\, \nabla I(x_i + u(x_i)), \end{aligned}$$

$$\begin{aligned} \frac{\partial \mathcal{E}}{\partial L_i} &= 2\alpha_1 L_i + 2\alpha_2 \sum_{j=1,\cdots,N, x_i \sim x_j} (L_i - L_j) \\ &+ \frac{\pi_0(J(x_i)|I^{\mathbf{b}}(x_i); \theta_0) - \pi_1(J(x_i)|I^{\mathbf{b}}(x_i); \theta_1)}{\pi(J(x_i)|I^{\mathbf{b}}(x_i); \theta)}, \\ \frac{\partial \mathcal{E}}{\partial \mu_1} &= -\frac{1}{\sigma_1^2} \sum_{i=1}^{N} p_i^1 \left(J(x_i) - I^{\mathbf{b}}(x_i) - \mu_1\right), \\ \frac{\partial \mathcal{E}}{\partial \sigma_1} &= -\frac{1}{\sigma_1^2} \sum_{i=1}^{N} p_i^1 \left(\frac{(J(x_i) - I^{\mathbf{b}}(x_i) - \mu_1)^2}{\sigma_1^2} - 1 \right), \\ p_i^0 &= \frac{(1 - L_i)\, \pi_0(J(x_i)|I^{\mathbf{b}}(x_i); \theta_0)}{\pi(J(x_i)|I^{\mathbf{b}}(x_i), L_i; \theta)}, \\ p_i^1 &= \frac{L_i\, \pi_1(J(x_i)|I^{\mathbf{b}}(x_i); \theta_1)}{\pi(J(x_i)|I^{\mathbf{b}}(x_i), L_i; \theta)}. \end{aligned} \tag{22}$$

The local weight p_i^j represents the probability that the pixel x_i may be considered as an element of the class j. It combines both a spatial information provided by L_i and an intensity information given by the probability ratio of π_j over π.

Besides, we use the projected gradient algorithm [59] to comply with the constraint on the parameter $\theta_1 = (\mu_1, \sigma_1)$: if $\mu_1^{(t+1)}$ is larger than μ_{max} (respectively smaller than μ_{min}), we replace it by μ_{max} (respectively μ_{min}) and we apply the same method for σ_1.

C. Simulating elastic deformations

In order to have a validation of the proposed classifying registration model, we will need to make experiments on simulated image pairs (see the experimental section). To obtain such pairs, we have to be able to simulate deformations that are samples from the probability distribution $\pi(\mathbf{b}) = \exp(-H(\mathbf{b}))$. The aim of this subsection is to explain how these samples are obtained. We consider on Ω_h the following neighborhood system: two different vertices e_i and e_j of Ω_h are neighbors ($e_i \sim e_j$) if they are on a same triangle. We denote by $\mathcal{C}_1$ and $\mathcal{C}_2$ the sets of cliques of order 1 and 2. From the decomposition (20), we can write

$$H(\mathbf{b}) \quad = \quad \underbrace{\frac{1}{2}\sum_{i=1}^{n_e} \beta_i^T A_{ii} \beta_i}_{\text{Sum on } \mathcal{C}_1} \; + \; \underbrace{\sum_{i,j=1, e_i \sim e_j}^{n_e} \beta_i^T A_{ij} \beta_j}_{\text{Sum on } \mathcal{C}_2} . \tag{23}$$

So, the elasticity displacement u is a Gibbs random field and, consequently, a Markov random field [60]. We can show that local conditional probability distributions $\pi(\beta_i | \beta_j, j \in \mathcal{N}_i)$ are Gaussian distributions with mean $-A_{ii}^{-1} W_i$, where $W_i = \sum_{j=1, j \neq i}^{n_e} A_{ij} \beta_i$, and covariance matrix A_{ii}^{-1}. The mean balances the displacement induced by neighboring vertices and the matrix A_{ii}^{-1} acts as a local filter. The vector $\mathbf{b}$ is a centered Gaussian Markovian random field. Elastic deformation can then be simulated using the Cholesky factorization of the covariance matrix A^{-1} [60].

D. Implementation details

For the implementation of the gradient descent algorithm, we use the FREEFEM++ software [61], dedicated to the finite element method. For the computational cost, we observed that, on images of size 200×200 pixels and with 500 finite elements, the algorithm approximately takes about one minute (the computations were performed on a core 2 duo 2GH and 2Go RAM processor).

In all the experiments, the initialization is always the same: $\mathbf{L} = 0$ and $u = 0$ (*i.e.* all pixels belong to a same first class and there is no displacement). In order to accelerate the convergence and the precision of the technique, we use a multi-grid approach, which consists in first initializing the deformations with a coarse approximation using a small number of vertices and then increasing this number when necessary. For an image of size 256×256, we use four levels of resolution with respectively 152, 302, 452, and 652 vertices. At each resolution, we have made four iterations of the gradient descent to estimate the coordinate of the displacement at each vertex.

IV. EXPERIMENTS

In this section, we show several experiments with the classifying registration model, both on synthetic and on real data. We compare this model to a classical one based on the sum of squared differences (SSD) and corresponding to a particular case of the classifying registration model obtained when the so-called class map $\mathbf{L}$ is set to be equal to 0 at every pixel. In all the experiments, we use a Gaussian distribution to describe the image intensity dependency law π_0 on the first class with parameter values (mean and variance) estimated from an image patch.

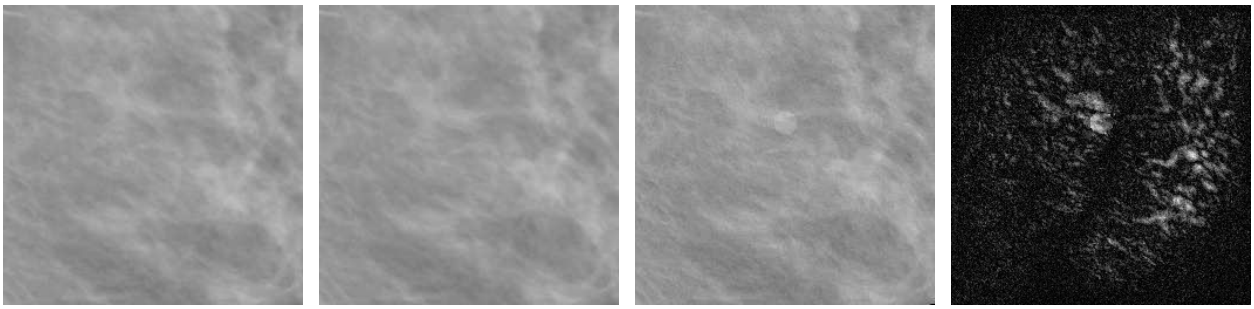
A. Model validation

Validation and comparison of medical image registration algorithms are two important issues [62], [63]. The requirements to perform a validation are: the definition of a validation methodology, which should include the design of validation datasets, the definition of a corresponding "ground truth", a validation protocol and some validation metrics. Now, there is rarely if ever a "ground truth" correspondence map that would enable judging the performance of a non-rigid registration algorithm. In our approach, we will use simulated deformations and images to get a "ground truth" to get an evaluation of registration and classification algorithms.

Different sub-images (of size 256×256 pixels each) were extracted from clinical screen-film mammograms of the MIAS database [64]. Each image was transformed by a simulated elastic deformation and corrupted by Gaussian noise with distribution $\mathcal{N}(0, 9)$. Mammographic masses were then simulated by adding to the deformed and noisy mammograms a Gaussian noise with distribution $\mathcal{N}(\mu_1, 9)$ on a disk-shaped region of radius R. Figure 2 shows a typical mammogram sub-image and its deformation including an additive mass.

From each mammogram sub-image, we have simulated several target images by changing parameter specifications. We thus obtained a database containing 576 pairs of images by varying lesion location, size $R \in \{0, 5, 10, 15\}$, contrast $\mu_1 \in \{5, 10, 15\}$ and deformation magnitude $D_m \in \{2, 4, 6\}$ (this deformation magnitude is defined as $D_m^2 = \frac{1}{N} \sum_{i=1}^{N} \parallel u(x_i) \parallel^2$). This database also contains 144 "normal" pairs of images without simulated mass.

We perform the registration of each pair of the database with both the full classifying model (Equation (17)) and the SSD model. For the classifying registration, we use a Gaussian mixture where π_0 is a Gaussian distribution $\mathcal{N}(0, 9)$ while the parameters (μ_1, σ_1) of π_1 are all estimated. For the regularization of the lesion map, we use a Bernoulli model (binary lesion map) with empirical choices $\alpha_1 = 7$ and $\alpha_2 = -0.7$.



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

Fig. 2. A simulated pair of images: (a) The source image extracted from a real mammogram, (b) The deformed source image, (c) The target image obtained by adding a Gaussian noise with distribution $\mathcal{N}(\mu_1, 9)$ on a disk-shaped region of the image (b), (d) Differences between the source and the target images.

We use two criteria for the evaluation of the registration. The first one involves the ground truth and compares the real displacement (u_{real}) and the estimated (u_{estim}) one. The registration error is defined as the mean distance between them, counted in pixels:

$$ErrL2 = \sqrt{\frac{1}{N}\sum_{i=1}^{N} ||u_{\text{real}}(x_i) - u_{\text{estim}}(x_i)||^2} \tag{24}$$

The second criterion is computed as the percentage of reduced differences between the source and target images:

$$DiffImg = 100 \cdot \left(1 - \frac{\sqrt{\sum_{i=1}^{N}(J(x_i) - I^{\mathbf{b}}(x_i))^2}}{\sqrt{\sum_{i=1}^{N}(J(x_i) - I(x_i))^2}}\right) \tag{25}$$

When the values of the criterion $DiffImg$ are negative, the $L2$ distance between the registered images is higher than the one between unregistered images. When it is computed only over lesion pixels, the second criterion can also be used for the evaluation of the detection. Negative values occuring on lesion regions show that the model preserves and enhances the image differences on points of the lesion class and thus improves lesion detection.

For a given pair of images we compute the values of $ErrL2$ on the output of both the classifying model and the SSD model. We also consider the difference of the two values thus obtained. The same computations are performed for the values of $DiffImg$. We show the statistical analysis of the obtained values by giving confidence intervals for confidence level of 0.6 [65].

Tables I and II summarize the obtained results, each line corresponding to a specific value of one of the parameters (μ_1, R or D_m); the confidence interval is computed over the pairs of the database having this specific parameter value.

	SSD model		Classifying model		Difference between both models	
	Global	Lesion	Global	Lesion	Global	Lesion
	$0.51\,;0.69$	$0.91\,;2.56$	$0.47\,;0.61$	$0.31\,;1.28$	$-0.09\,;0.02$	$-1.34\,;0.39$
$R=5$	$0.51\,;0.66$	$0.55\,;1.70$	$0.45\,;0.60$	$0.30\,;1.04$	$-0.01\,;0.01$	$-1.03\,;0.13$
10	$0.52\,;0.69$	$0.93\,;2.56$	$0.45\,;0.61$	$0.36\,;1.37$	$-0.06\,;0.02$	$-1.73\,;0.15$
15	$0.53\,;0.74$	$0.90\,;2.69$	$0.44\,;0.63$	$0.41\,;1.87$	$-0.16\,;0.043$	$-1.35\,;0.92$
$\mu_1=10$	$0.46\,;0.62$	$0.74\,;1.72$	$0.45\,;0.62$	$0.50\,;1.96$	$-0.01\,;0.02$	$-0.30\,;0.38$
15	$0.50\,;0.67$	$0.95\,;2.41$	$0.45\,;0.61$	$0.34\,;1.10$	$-0.06\,;0.01$	$-1.35\,;-0.26$
20	$0.50\,;0.70$	$1.08\,;2.92$	$0.45\,;0.60$	$0.42\,;0.91$	$-0.13\,;0.01$	$-2.09\,;-0.45$
$D_m=2$	$0.53\,;0.59$	$0.92\,;2.13$	$0.53\,;0.57$	$0.39\,;0.80$	$-0.02\,;0.00$	$-1.26\,;0.04$
4	$0.56\,;0.62$	$0.89\,;2.32$	$0.57\,;0.61$	$0.30\,;1.04$	$-0.02\,;0.01$	$-1.20\,;0.18$
6	$0.64\,;1.64$	$0.67\,;3.58$	$1.31\,;2.20$	$0.45\,;3.91$	$0.39\,;0.79$	$-1.38\,;1.14$

TABLE I

A FIRST EVALUATION OF THE REGISTRATION PERFORMANCES OF THE SSD MODEL AND OF THE CLASSIFYING MODEL. THIS TABLE PRESENTS THE CONFIDENCE INTERVALS OF LEVEL 0.6 OF THE MEAN DISTANCE CRITERION $ErrL2$ (EQUATION (24)) BETWEEN THE REAL AND THE ESTIMATED DEFORMATIONS, COMPUTED OVER THE WHOLE IMAGE (COLUMN "GLOBAL") OR ONLY ON THE "LESION". THE LAST DOUBLE-COLUMN GIVES THE CONFIDENCE INTERVAL OF THE DIFFERENCE BETWEEN VALUES OF THE $ErrL2$ CRITERION OBTAINED WITH THE SSD AND CLASSIFYING MODELS. A NEGATIVE VALUE OF THE DIFFERENCE INDICATES THAT THE CLASSIFYING MODEL IS MORE ACCURATE THAN THE SSD ONE.

Columns "global" of Table I reveal that the accuracy of the classifying model and of the SSD model are about the same when evaluated over whole images (*i.e.* inside and outside lesions); the differences between the $ErrL2$ values of both models vary between -0.09 and 0.02 pixels. However, on columns "lesion" of Table I, we observe that SSD registrations are impaired by the presence of a lesion whereas those of the classifying model seem to be more robust. The SSD model accuracy on lesions is lower than the global one, and decreases as the lesions become larger or more constrasted. This is in sharp contrast with the classifying model whose accuracy on lesions is close to the global one and does not vary significantly according to lesions size and constrast.

Table II gives another viewpoint on model evaluation, which completes the previous one. On columns "global", we can see that the SSD model reduces slightly more source and target intensity differences than the classifying one. On columns "lesion", we further observe that, on lesions, these differences are more compensated by the SSD model than by the classifying one. Contrarily to the SSD model, the classifying model can detect lesions and preserves the differences they cause. With the classifying model,

	SSD model		Classifying model		Difference between both models	
	Global	Lesion	Global	Lesion	Global	Lesion
	$35\ ;\ 51$	$5\ ;\ 31$	$30\ ;\ 43$	$-2\ ;\ 24.$	$-2\ ;\ 0$	$-14\ ;\ -4$
$R = 5$	$36\ ;\ 51$	$4\ ;\ 35$	$31\ ;\ 43$	$-11\ ;\ 23$	$-0.3\ ;\ 0$	$-10\ ;\ 0.6$
10	$39\ ;\ 55$	$4\ ;\ 32$	$30\ ;\ 42$	$2\ ;\ 26$	$-1\ ;0$	$-13\ ;\ -4$
15	$34\ ;\ 50$	$5\ ;\ 28$	$34\ ;\ 49$	$-3, 17$	$-3\ ;0$	$-15\ ;\ -7$
$\mu_1 = 10$	$36\ ;\ 52$	$10\ ;\ 42$	$31\ ;\ 43$	$-2\ ;\ 28$	$-1\ ;\ 0$	$-9\ ;\ -1$
15	$35\ ;\ 51$	$5\ ;\ 32$	$30\ ;\ 43$	$-1\ ;\ 24$	$-2\ ;\ 0$	$-16\ ;\ -7$
20	$36\ ;\ 51$	$6\ ;\ 27$	$28\ ;\ 43$	$-3\ ;\ 17$	$-3\ ;\ 0$	$-15\ ;\ -7$
$D_m = 2$	$30\ ;\ 34$	$5\ ;\ 26$	$30\ ;\ 34$	$-6\ ;\ 13$	$-1\ ;\ 0$	$-14\ ;\ -4$
4	$43\ ;\ 49$	$6\ ;\ 34$	$42\ ;\ 49$	$-1\ ;\ 28$	$-1\ ;\ 0$	$-13\ ;\ -4$
6	$48\ ;\ 55$	$5\ ;\ 36$	$41\ ;\ 50$	$-1\ ;\ 26$	$-7\ ;\ -4$	$-15\ ;\ -4$

TABLE II

A SECOND EVALUATION OF THE REGISTRATION PERFORMANCES OF THE SSD MODEL AND OF THE CLASSIFYING MODEL. THIS TABLE PRESENTS THE CONFIDENCE INTERVALS OF LEVEL 0.6 OF THE MEAN PERCENTAGE OF REDUCED DIFFERENCES CRITERION $DiffImg$ (EQUATION (25)) BETWEEN THE SOURCE AND THE TARGET IMAGES, COMPUTED OVER THE WHOLE IMAGES (COLUMN "GLOBAL") OR ONLY ON THE "LESIONS". THE LAST DOUBLE-COLUMN GIVES THE CONFIDENCE INTERVAL OF THE DIFFERENCES BETWEEN THE VALUES OF THE $DiffImg$ CRITERION OBTAINED WITH THE CLASSIFYING AND THE SSD MODELS.

lesions get better and better preserved in image differences when they get larger or more contrasted.

Detection evaluation: Lesion detection from bilateral or temporal mammograms can be performed by thresholding the differences between the registered images [66]. The classifying model provides a lesion map $\mathbf{L}$ whose spatial consistency is ensured by the use of a prior term. However, to compare the detection performances of the SSD model and of the classifying model depending on registration results, we threshold the registered image differences obtained with both models. When comparing the binary image obtained by thresholding to the ground truth, we can determine for each experiment the number of false positives (FP), false negatives (FN), true positives (TP) and true negatives (TN). A graphical representation of these two indices are obtained as a Receiver Operating Characteristic (ROC) curve [67] expressing the sensitivity ($Sensitivity = \frac{TP}{TP+FN}$) versus the false positive rate ($Specificity = \frac{FP}{TN+FP}$), for different values of the thr! eshold. In such a representation, the best possible prediction method would yield a point in the upper left corner (coordinate $(0, 1)$ of the ROC space) representing 100% sensitivity (no false negatives) and 100% specificity (no false positives). On Figure 3, we superimpose the ROC curves obtained with the two registration models. Figure 3 shows that the classifying registration method

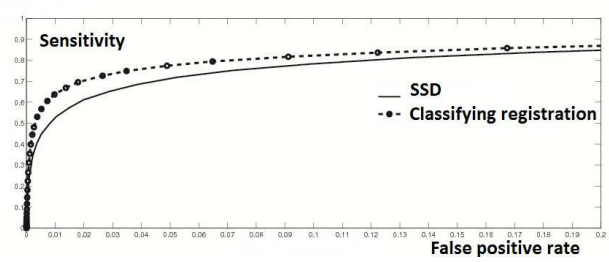


Fig. 3. ROC curves of the detection results obtained respectively with the SSD and classifying models.

improves the detection results by reducing the false positive rate. For instance, for a detection rate of 80% (resp. 70%), the false positive rate is 0.115% (resp. 0.042%) with the SSD model and only 0.064% (resp. 0.018%) with the classifying model.

B. Some other applications of the model

1) Template-based segmentation: In the example shown on Figure 4, we apply the classifying model to contrast-enhanced CT axial images of the thorax. The contrast agent enhances the superior vena cava (V) and the aorta arch (C). First, we manually segment these two regions in the post-contrast image. Then, we register the post-contrast image to the pre-contrast one by using either the SSD model or the classifying model with the exact class map and a uniform distribution π_1. Registering the post-contrast image (here taken as the source image) with the pre-contrast image (taken as the target image) also allows us to transport the segmentation made on the source image to the target image; such a technique is known as template-based segmentation. The results, shown on Figure 4, illustrate the superiority of the classifying model which, contrarily to the SSD model, takes into account the contrast variations and does not shrink contrast-enhanced regions. We further quantified the amount of shrinkage on the enhanced regions on 9 couples of images representing several different cases. On average, the amount of shrinkage was 29.5% for the SSD-based registration and only 6.35% for the classifying registration.

2) Mammogram registration: On the example shown on Figure 5, we process a pair of bilateral mammograms (case 201 of the MIAS database [64]) which contains a lesion appearing as an asymmetric density.

We apply the classifying model in the class map estimation mode (minimization of Equation (17)): the binary class map $\mathbf{L}$ is estimated, together with lesion-related parameters μ_1 and σ_1 of the distribution π_1. A Bernoulli prior is used for the definition of the binary lesion map with heuristic choices $\alpha_1 = 7$

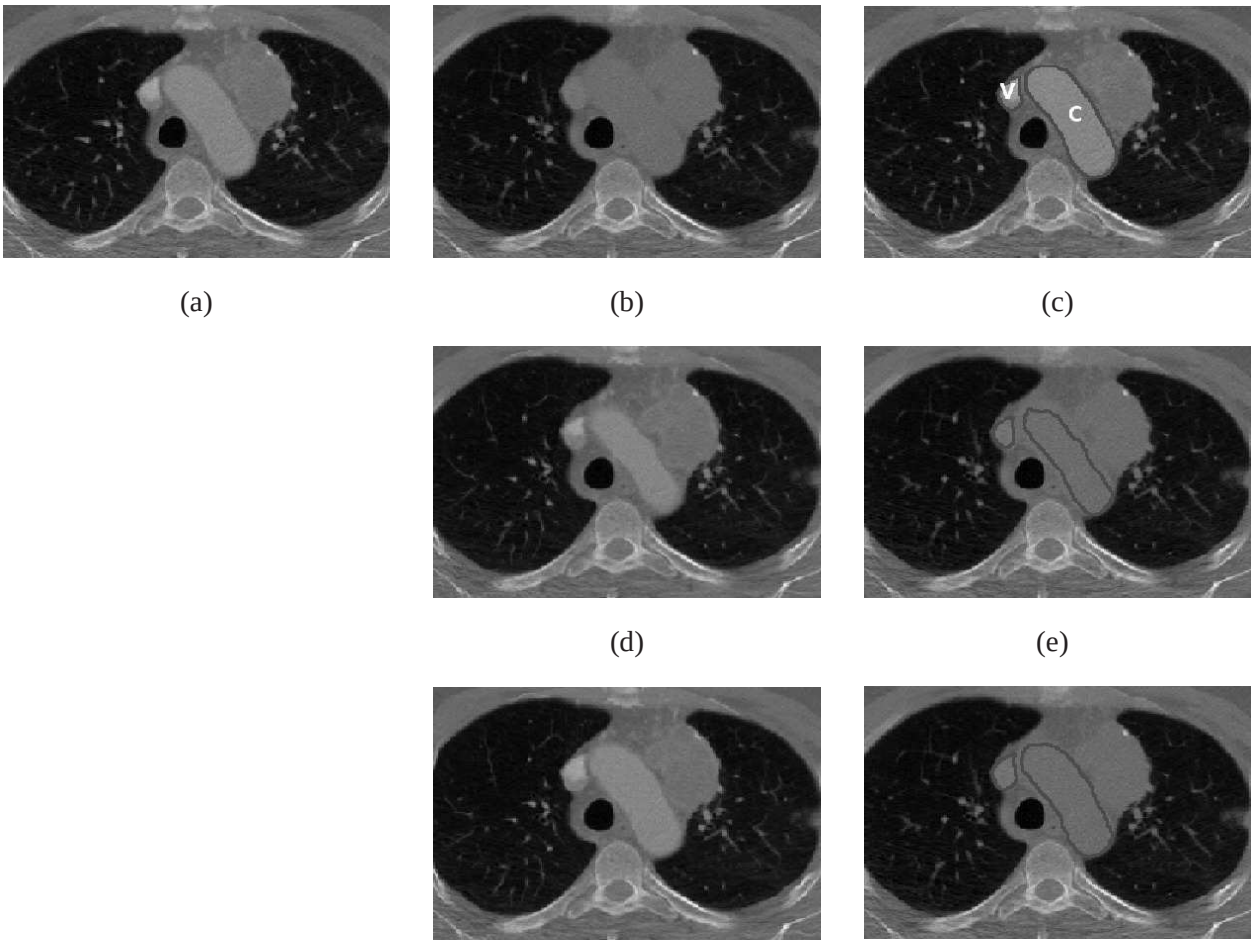


Fig. 4. A comparison of the SSD model and the classifying model for template-based segmentation of contrast-enhanced images. For the classifying model, the class map is known and the distribution π_1 is set as a uniform distribution. Image (a) is the post-contrast image (source image), image (b) is the pre-contrast image (target image), and image (c) is the two-class segmentation of the post-contrast image. Images (d) and (e) are respectively the registered source image and the transported segmentation obtained by the SSD registration, whereas images (f) and (g) are those obtained by the classifying model.

et $\alpha_2 = -0.7$.

By comparison of images (c) and (i) of Figure 5, we can see that the initial image differences inside the lesion region are preserved after a classifying registration while those outside are attenuated. In other words, the classifying model enhances the lesion in image differences. Such a feature is particularly interesting for lesion detection as it can help reducing the number of false positives. The classifying model also overcomes a drawback of the usual SSD model which tends to reduce both normal and pathological asymmetries appearing in image differences, as it can be observed on image (f) of Figure 5.

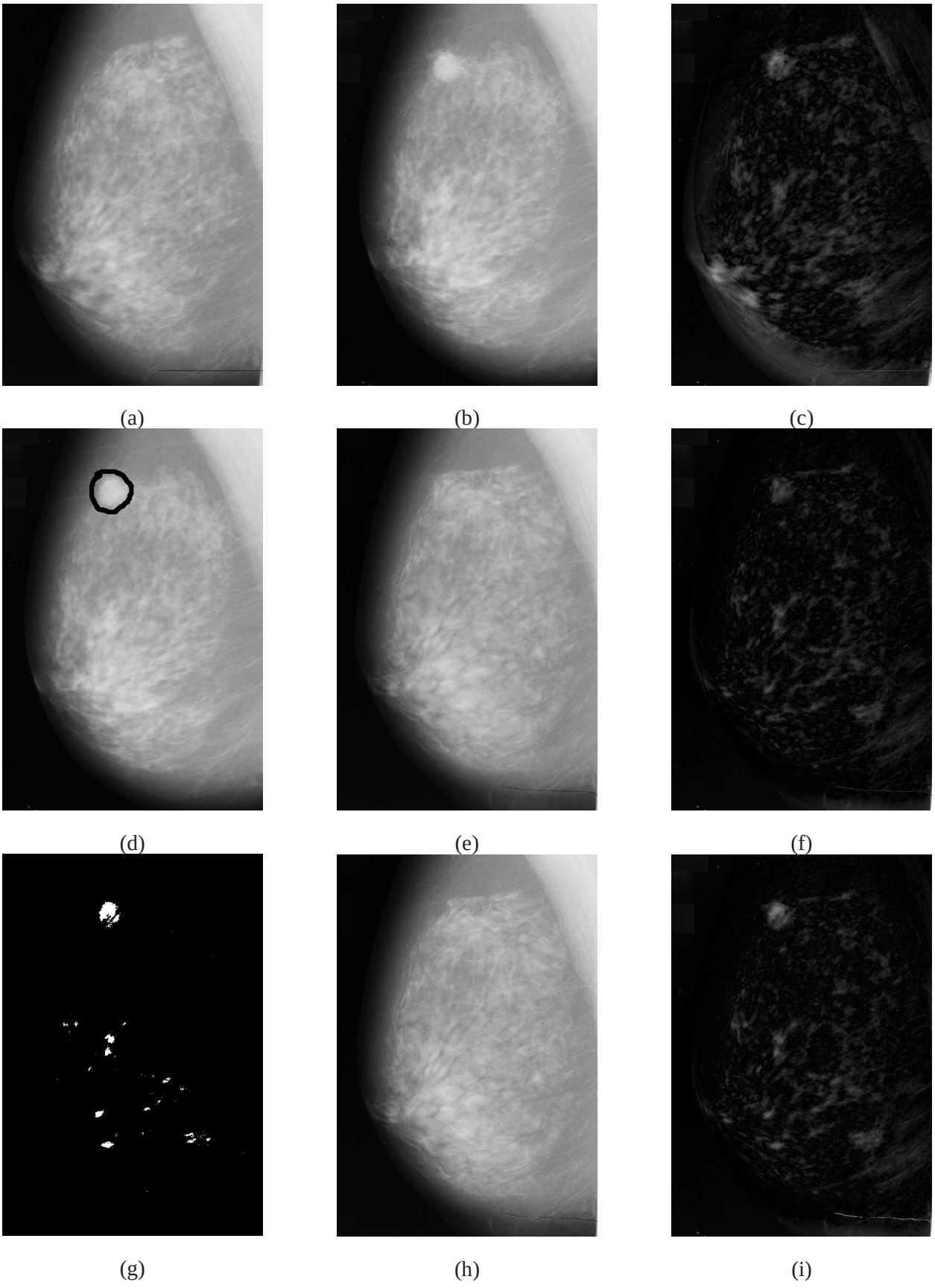


Fig. 5. Images (a) and (b) form a bilateral pair of mammograms showing a pathological asymmetry contoured in image (d). Image (c) is the initial difference between images (a) and (b). Image (e) is the deformation of image (a) obtained by the SDD registration between images (a) and (b). Image (f) is the difference between image (b) and the registered image (e). The last row shows the results of the classifying registration of images (a) and (b) : image (g) is the obtained class map L ; image (h)

V. Discussion

In this paper, we have proposed a Bayesian approach to perform simultaneously image registration and pixel classification.

The proposed technique is well-suited to deal with image pairs that contain two classes of pixels with different inter-image intensity relationships. We have shown through different experiments that the model can be applied in many different ways. For instance if the class map is known, then it can be used for template-based segmentation. If the full model is used (estimation of the class map, the registration and the parameters of the distribution of the outliers), then it can be applied to lesion detection by image comparison.

Experiments have been conducted on both real and simulated data. They show that in the presence of an extra-class (e.g. a lesion class in mammograms), the classifying registration improves both the registration and the detection, especially when the deformations are small. The proposed model is defined using only two classes but it is straightforward to extend it to an arbitrary number of classes. However, the estimation of the number of classes would then appear as a critical issue. This will be part of some future research and it will certainly require the use of model selection techniques.

The application of the classifying model was illustrated on medical imaging data. But, the proposed model is very generic and can be adapted to many other situations. In particular, we believe that the model could also be helpful for motion estimation. The introduction of a second intensity relationship class in the model would enable to deal with occlusions, which are a major issue of motion estimation.

References

[1] L. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, vol. 24, pp. 325–376, 1992.

[2] D. Hill, P. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," *Physics in Medicine and Biology*, vol. 46, no. 3, pp. R1–R45, 2001.

[3] J. Maintz and M. Viergever, "A survey of medical image registration," *Medical Image Analysis*, vol. 2(1), pp. 1–36, 1998.

[4] P. Pluim, J. Maintz, and M. Viergever, "Mutual-information based registration of medical images : a survey," *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 986–1004, 2003.

[5] J. Modersitzki, *Numerical Methods for Image Registration*, ser. Numerical Mathematics And Scientific Computation, C. S. G. H. Golub and E. Suli, Eds. Oxford University Press, 2004.

[6] R. Woods, J. Mazziotta, and S. R. Cherry, "MRI-PET registration with automate algorithm," *Journal of Computer Assisted Tomography*, vol. 17, no. 4, pp. 536–46, Jul-Aug 1993.

[7] T. M. Buzug and J. Weese, "Voxel-based similarity measures for medical image registration in radiological diagnosis and image guided surgery," *Journal of Computing and Information Technology*, vol. 6, no. 2, pp. 165–179, 1998.

[8] A. Roche, G. Malandain, X. Pennec, and N. Ayache, "The correlation ratio as a new similarity measure for multimodal image registration," in *First Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI'98)*, vol. LNCS 1496. Cambridge, USA: Springer Verlag, October 1998, pp. 1115–1124.

[9] C. Studholme, D. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognition*, vol. 1, no. 32, pp. 71–86, 1999.

[10] A. Roche, G. Malandain, and N. Ayache, "Unifying Maximum Likelihood Approaches in Medical Image Registration," *International Journal of Computer Vision of Imaging Systems and Technology*, vol. 11, pp. 71–80, 2000.

[11] G. Malandain, "Les mesures de similarité pour le recalage des images médicales," Habilitation à Diriger des Recherches, Université de Nice Sophia-Antipolis, 2006.

[12] P. Hayton, M. Brady, L. Tarassenko, and N. Moore, "Analysis of dynamic MR breast images using a model of contrast enhancement," *Medical Image Analysis*, vol. 1, pp. 207–224, 1997.

[13] C. Tanner, J. A. Schnabel, D. Chung, M. J. Clarkson, D. Rueckert, D. Hill, and D. J. Hawkes, "Volume and shape preservation of enhancing lesions when applying non-rigid registration to a time series of contrast enhancing MR breast images," in *MICCAI '00: Proceedings of the Third International Conference on Medical Image Computing and Computer-Assisted Intervention*, London, UK, 2000, pp. 327–337.

[14] T. Rohlfing and C. R. Maurer, Jr., "Intensity-based non-rigid registration using adaptive multilevel free-form deformation with an incompressibility constraint," in *Proceedings of Fourth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2001)*, ser. Lecture Notes in Computer Science, W. Niessen and M. A. Viergever, Eds., vol. 2208. Berlin: Springer-Verlag, 2001, pp. 111–119.

[15] N. Sebe and M. Lew, *Robust Computer Vision: Theory and Applications*. Series: Computational Imaging and Vision, 2003, vol. 26.

[16] P. Rousseeuw and A. Leroy, *Robust Regression and Outlier Detection*. Wiley, New York, 1987.

[17] F. Hampel, E. Ronchetti, P. Rousseeuw, and W. Stahel, *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York, 1986.

[18] D. Hawkins, *Identifications of Outliers*. Chapman & Hall, London, New York, 1980.

[19] P. Huber, *Robust Statistics*. Wiley, New York, 1981.

[20] F. Richard, "A new approach for the registration of images with inconsistent differences," in *Proc. of the Int. Conf. on Pattern Recognition, ICPR*, vol. 4, Cambridge, UK, 2004, pp. 649–652.

[21] C. Nikou, F. Heitz, and J.-P. Armspach, "Robust voxel similarity metrics for the registration of dissimilar single and multimodal images," *Pattern Recognition*, vol. 32, pp. 1351–1368, 1999.

[22] J. Kim, "Intensity based image registration using robust similarity measure and constrained optimization: Applications for radiation therapy," Ph.D. dissertation, University of Michigan, 2004.

[23] M. Black, "Robust incremental optical flow," Ph.D. dissertation, Yale University, New Haven, CT, USA, 1992.

[24] P. Meer, D. Mintz, A. Rosenfeld, and D. Kim, "Robust regression methods for computer vision: A review," *International Journal of Computer Vision*, vol. 6, no. 1, pp. 59–70, 1991.

[25] M. Black and A. Rangarajan, "On the unification of line processes, outlier rejection, and robust statistics with applications in early vision," *International Journal of Computer Vision*, vol. 19, no. 1, pp. 57–91, 1996.

[26] N. S. Netanyahu and I. Weiss, "Analytic outlier removal in line fitting," in *12th IAPR International Conference on Computer Vision and Image Processing*, vol. 2B, 1994, pp. 406–408.

[27] P. Schroeter, J.-M. Vesin, T. Langenberger, and R. Meuli, "Robust parameter estimation of intensity distributions for brain magnetic resonance images," *IEEE Transactions on Medical Imaging*, vol. 17, no. 2, pp. 172–186, 1998.

[28] P. S. Torr, R. Szeliski, and P. Anandan, "An Integrated Bayesian Approach to Layer Extraction from Image Sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, pp. 297–303, 2001.

[29] D. Hasler, L. Sbaiz, S. Susstrunk, and M. Vetterli, "Outlier modeling in image matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 301–315, 2003.

[30] P. Biber, S. Fleck, and W. Straßer, "A probabilistic framework for robust and accurate matching of point clouds," in *26th Pattern Recognition Symposium (DAGM 04)*, 2004.

[31] A. Jepson and M. Black, "Mixture models for optical flow computation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 1993, pp. 760–761.

[32] M. Hachama, A. Desolneux, and F. Richard, "Combining registration and abnormality detection in mammography," in *Workshop on Biomedical Image Registration, (WBIR'2006)*, ser. Lecture Notes in Computer Science, J. Pluim, B. Likar, and F. Gerritsen, Eds., vol. 4057. Springer, 2006, pp. 178–185.

[33] H. Hachama, F. Richard, and A. Desolneux, "A mammogram registration technique dealing with outliers," in *Proc. of the IEEE International Symposium on Biomedial Imaging (ISBI'2006)*, 2006, pp. 458–461.

[34] M. Hachama, F. Richard, and A. Desolneux, "A probabilistic approach for the simultaneous mammogram registration and abnormality detection," in *International Workshop on Digital Mammography (IWDM'2006)*, ser. Lecture Notes in Computer Science, S. Astley, M. Brady, C. Rose, and R. Zwiggelaar, Eds., vol. 4046. Springer, 2006, pp. 205–212.

[35] N. Rougon, A. Discher, and F. Preteux, "Region-driven statistical non-rigid registration: application to model-based segmentation and tracking of the heart in perfusion MRI," in *Proceedings SPIE Conference on Mathematical Methods in Pattern and Image Analysis*, vol. 5916, San Diego, August 2005, pp. 148–159.

[36] C. Studholme, D. Hill, and D. J. Hawkes, Eds., *Incorporating Connected Region Labeling into Automated Image Registration Using Mutual Information*, ser. Proceedings of the IEEE workshop on Mathematical Methods in Biomedical Image Analysis, San Francisco Ca. IEEE Computer Society Press, 1996.

[37] J. Chappelow, B. Bloch, N. Rofsky, E. Genega, R. Lenkinski, W. DeWolf, and A. Madabhushi, "Elastic registration of multimodal prostate mri and histology via multiattribute combined mutual information." *Med Phys*, vol. 38, no. 4, pp. 2005–18, 2011.

[38] P. Patel, J. Chappelow, J. Tomaszewski, M. Feldman, M. Rosen, N. Shih, and A. Madabhushi, "Spatially weighted mutual information (SWMI) for registration of digitally reconstructed ex vivo whole mount histology and in vivo prostate MRI," in *IEEE International Conference of Engineering in Medicine and Biology Society (EMBS)*, 30 2011-sept. 3 2011, pp. 6269 –6272.

[39] M. Hachama, A. Desolneux, C. Cuenod, and F. Richard, "A classifying registration technique for the estimation of enhancement curves of DCE-CT scan sequences," *Medical Image Analysis*, vol. 14, no. 2, pp. 185 – 194, 2010.

[40] J. Ashburner and K. Friston, "Unified segmentation," *NeuroImaging*, vol. 26, pp. 839–851, 2005.

[41] K. Pohl, J. Fisher, W. Grimson, R. Kikinis, and W. Wells, "A Bayesian model for joint segmentation and registration," *NeuroImage*, vol. 31, no. 1, pp. 228–239, 2006.

[42] P. Wyatt and A. Noble, "MAP MRF Joint Segmentation and Registration of Medical Images," *Medical Image Analysis*, vol. 7, no. 4, pp. 539–552, 2003.

[43] C. Xiaohua, M. Brady, J. L.-C. Lo, and N. Moore, "Simultaneous segmentation and registration of contrast-enhanced breast MRI," in *Information Processing in Medical Imaging*, 2005, pp. 126–137.

[44] Y. Zheng, J. Yu, C. Kambhamettu, S. Englander, M. Schnall, and D. Shen, "De-enhancing the Dynamic Contrast-Enhanced Breast MRI for Robust Registration," in *MICCAI*, ser. Lecture Notes in Computer Science, N. Ayache, S. Ourselin, and A. Maeder, Eds., vol. 4791. Springer, 2007, pp. 933–941.

[45] D. Mumford, "The Bayesian rationale for energy functionals," *Geometry-driven Diffusion in Computer Vision*, vol. 46, pp. 141–153, 1994.

[46] C. A. Glasbey and K. V. Mardia, "A review of image-warping methods," *Journal of Applied Statistics*, vol. 25, no. 2, pp. 155–171, April 1998.

[47] C. Broit, "Optimal registration of deformed images," Ph.D. dissertation, University of Pennsylvania, Philadelphia, 1981.

[48] R. Bajcsy, R. Lieberson, and M. Reivich, "Computerized system for the elastic matching of deformed radiographic images to idealized atlas images," *Journal of Computer Assisted Tomography*, vol. 7, 1983.

[49] R. Bajcsy and S. Kovacic, "Multiresolution elastic matching," *Computer Vision, Graphics, and Image Processing*, vol. 46, no. 1, pp. 1–21, 1989.

[50] M. Miller, G. Christensen, Y. Amit, and U. Grenander, "Mathematical textbook of deformable neuroanatomies," *Proc. of the National Academy of Sciences*, vol. 90, pp. 11 944–11 948, 1993.

[51] Y. Amit, U. Grenander, and M. Piccioni, "Structural image restoration through deformable templates," *American Statistical Association*, vol. 86, no. 414, June 1991.

[52] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 6, pp. 567–585, 1989.

[53] S. Allassonnière, Y. Amit, and A. Trouvé, "Towards a coherent statistical framework for dense deformable template estimation," *Journal Of The Royal Statistical Society Series B*, vol. 69, no. 1, pp. 3–29, 2007.

[54] F. Richard, A. Samson, and C. Cuenod, "A SAEM algorithm for the estimation of template and deformation parameters in medical image sequences," *Statistics and Computing*, vol. 19, no. 4, pp. 465–478, 2009.

[55] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721–741, 1984.

[56] B. Chalmond, *Modelling and inverse problems in image analysis*, ser. Applied mathematical sciences, Kindle, Ed. Springer, 2003, vol. 155.

[57] M. Staring, S. Klein, and J. Pluim, "Nonrigid registration with tissue-dependent filtering of the deformation field," *Physics in Medicine and Biology*, vol. 52, pp. 6879–6892, 2007.

[58] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*. Amsterdam: North-Holland Publishing Co., 1978.

[59] J. Bonnans, J. Gilbert, C. Lemaréchal, and C. Sagastizábal, *Numerical Optimization – Theoretical and Practical Aspects*, ser. Universitext. Springer Verlag, Berlin, 2006.

[60] H. Rue and L. Held, *Gaussian Markov Random Fields: Theory and Applications*, ser. Monographs on Statistics and Applied Probability. London: Chapman & Hall, 2005, vol. 104.

[61] F. Hechtand, O. Pironneau, A. L. Hyaric, and K. Ohtsuka, "Freefem++ manual," *www.freefem.org/*, October 2005.

[62] R. Woods, "Validation of registration accuracy," in *Handbook of medical imaging*, I. Bankman, Ed. Orlando, FL, USA: Academic Press, Inc., 2000, pp. 491–497.

[63] P. Hellier, C. Barillot, I. Corouge, B. Gibaud, G. L. Goualher, D. Collins, A. Evans, G. Malandain, N. Ayache, G. Christensen, and H. Johnson, "Retrospective evaluation of inter-subject brain registration," *IEEE Transactions on Medical Imaging*, vol. 22, no. 9, pp. 1120–1130, 2003.

[64] J. Suckling, J. Parker, D. Dance, S. Astley, and et al., "The mammographie image analysis society digital mammogram database," in *Proceedings of the 2nd international workshop on digital mammography*, 1994.

[65] T. Soong, *Fundamentals Of Probability and Statistics for Engineers*. State University of New York at Buffalo, USA: John Wiley and Sons, 2004.

[66] M. Sallam and K. Bowyer, "Registration and difference analysis of corresponding mammogram images," *Medical Image Analysis*, vol. 3, no. 2, pp. 103–118, 1999.

[67] C. Metz, "ROC Methodology in Radiologic Imaging," *Investigative Radiology*, vol. 21, pp. 720–733, 1986.

Mohamed Hachama received the Master Degree in "mathematics, vision and learning (MVA)" from ENS Cachan (2004), a PhD degree from the University Paris Descartes (2008) and the Algerian Habilitation (2011) from the Saad Dahlab University. He is currently an associate professor in applied mathematics at Khemis Miliana University. He works on variational and stochastic models for image processing and their applications to medical images (image registration and classification,3D surface analysis, etc.)

Agnès Desolneux is currently researcher for the french National Center of Scientific Research (CNRS) in the applied mathematics department of the Ecole Normale Supérieure de Cachan. She has studied mathematics at the Ecole Normale Superieure in Paris and she has received her PhD degree in 2000. She works on stochastic models for image analysis and she has co-authored two books: one with Lionel Moisan and Jean-Michel Morel entitled From Gestalt theory to image analysis: a probabilistic approach (Springer, 2008), and one with David Mumford entitled Pattern Theory: The stochastic analysis of real-world signals (AK Peters, 2010).

Frédéric Richard is currently professor in applied mathematics at Aix-Marseille University. After studies in the master "mathematics, vision, and learning" of the Ecole Normale Supérieure of Cachan, he received a PhD degree (2001) and the French habilitation (2009) from the University Paris Descartes. His research activities are mainly focused on statistical aspects of image processing. He published works about different image processing issues (texture analysis, image registration, 3D surface analysis, etc.) and their applications to biomedical images from various modalities (radiography, contrast-enhanced imaging, IRPE, etc.).