INTERNATIONAL JOURNAL OF
NUMERICAL ANALYSIS AND MODELING
Volume 6, Number 2, Pages 311–334

A ROBUST AFFINE IMAGE REGISTRATION METHOD

NOPPADOL CHUMCHOB AND KE CHEN

Abstract. Image registration has many real life applications. Affine image registration is one of the commonly-used parametric models. Iterative solution methods for the underlying least squares problem suffer from convergence problems whenever good initial guesses are not available. Variational models are non-parametric deformable models that have been proposed based on least squares fitting and regularization. The fast iterative solution methods often require a reliable parametric (affine) method in a pre-registration step. In this paper we first survey and study a class of methods suitable for providing the good initial guesses for the affine model and a diffusion based variational model. It appears that these initialization methods, while useful for many cases, are not always reliable. Then we propose a regularized affine least squares approach that can overcome the convergence problems associated with existing methods. Combined with a cooling idea in a multiresolution setting, it can ensure robustness and selection of the optimal coupling parameter efficiently. Numerical examples are given to demonstrate the effectiveness of our proposed approach.

Key Words. Image registration, affine transformation, regularization, Newton method.

1. Introduction

Image registration is the process of spatially aligning two or more images of the same object taken in different times or from different viewpoints or by different imaging machineries as in multi-modality imaging. When two images are taken as input, one of them is called the reference image and is kept unchanged and used as the reference, whereas the other is called the template image and is employed to register the reference image. The goal of image registration is to determine an optimal transformation in such a way that the transformed template image becomes *similar* to the reference image as much as possible, mapping points from the template image onto the reference image. This transformation is sought from optimizing an appropriate object functional, which measures the similarity of the transformed template image to the reference image and the transformation regularity; it is the latter that we address here in a new model context.

The registration methodology can be classified into two main *physical* categories: rigid and non-rigid registration, or *mathematical* categories: linear and nonlinear registration, or *complexity* categories: parametric and non-parametric registration.

Received by the editors January 1, 2008 and, in revised form, March 22, 2008.
2000 *Mathematics Subject Classification.* 62B5, 94A08, 47A52, 90C53.
This research was supported by a Royal Thai Government Fellowship for NC. For correspondence, contact k.chen@liv.ac.uk.

On one hand, rigid registration involves a linear rigid-body transformation, consisting of rotation and translation (with only 3 unknown parameters). On the other hand, deformable registration (e.g. variational models [23]) may include nonlinear transformations (non-parametric) whose number of unknowns for a discrete image is proportional to the number of pixels!

In real-life applications, rigid registration alone cannot always provide a satisfactory result, particularly in many medical applications (e.g. one cannot ensure the patient sits in the identical position with respect to the equipment each time), while deformable registration may not be quick enough for ready use. In this paper, we are mainly concerned with affine transformation for 2 reasons: (i) it is applicable to a large class of non-rigid registration problems. Moreover, as our experience shows, an affine method is always many orders of magnitude faster than a nonlinear variational method [23] due to much less unknowns involved. See also Thévenaz *et al.*[32], Jenkinson and Smith [15], Jenkinson *et al.*[14], Modersitzki [23], Xia and Liu [36], Zhilkin and Alexander [38], Lucchese [20] and the references therein. (ii) it is widely used as a pre-registration step for sophistical non-rigid registration methods, such as elastic, fluid, and diffusion registration, by providing the good initial positions for the image to be registered (see [23] and Schmitt *et al.*[29] and the references therein).

To realize image registration, assuming the image intensities of the given images are comparable, a common approach is to minimise the Sum of Squared Differences (SSD) i.e. the least-squared function for the squared pixel-wise differences in the image intensity between the transformed template image and the reference image. Then the obtained transformation defines a transformed template image as required. For affine registration, although there are only 6 parameters, iterative methods to solve the underling nonlinear minimization can suffer from convergence problems if good initial guesses are not possible (i.e. even after we attempt to devise good initial guesses). A theoretical reason may be that image registration problem is ill-posed in the sense of Hadamard. The information provided by the reference and the least-squared model are not sufficient to ensure the existence, uniqueness, and stability of a solution [12]. This motivates us to introduce regularization into affine registration, as one would do for other ill-posed problems [1, 23, 26, 34]. The result is a refined affine registration model that can be solved by converging methods for a large class of image problems.

We remark that image registration is required whenever comparing a series of images is of interest. For example, in remote sensing applications, registration of satellite images taken over a region during different seasons or years can be used to detect environment change over time [4, 5, 27] while, in medical image processing, a vital component of applications is the registration of relevant images of a patient in order to obtain accurate information for diagnosis, monitoring disease progression, planing treatment and treatment guidance. See Maintz and Viergever [22], Hajnal *et al.* [9], and Hill *et al.* [13].

The rest of the paper is organized as follows. We introduce the affine and the diffusion image registration respectively in Sections 2 and 3, and then present four methods to improve affine registration in Section 4. A regularized affine registration (RAR) model is presented in Section 5 followed by a regularization parameter selection algorithm in Section 6. Some numerical experiments on the performance of the proposed method are presented in Section 7, followed by conclusions in Section 8.

2. The preliminaries, affine image registration and solution methods

Assume that in continuous variables the given images can be represented by compactly supported functions $R, T : \Omega \subset \mathbb{R}^2 \rightarrow V \subset \mathbb{R}_0^+$. It is customary to consider $\Omega = [0,1]^2$ and $V = [0,1]$ for gray-scale images. In practice, two discrete images of the same size $n_1 \times n_2$ are given: the reference R and the template T.

For each pixel $\mathbf{x} = (x_1, x_2)^\top$, denote by $\varphi = \varphi(\mathbf{x}) : \Omega \rightarrow \Omega$ the *unknown* coordinate transformation that produces the alignment between the reference R and the transformed version of the template

$$F = T \circ \varphi = T_\varphi(\mathbf{x}) = T(\varphi(\mathbf{x})). \tag{1}$$

We hope to achieve that $F \approx R$ or $F - R \approx 0$. Here the transformation φ has 2 components

$$\varphi(\mathbf{x}) = (\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}))^\top. \tag{2}$$

Once $\varphi(\mathbf{x})$ for each $\mathbf{x}$ is calculated, an interpolation scheme is required to assign the image intensity values for the transformed template image F at possibly non-grid location $\varphi(\mathbf{x})$ within image boundaries. For regions outside the image boundaries, the image intensities can be set to be a constant value, usually zero.

All registration strategies require a suitable *similarity measure* $\mathcal{D}$ (also called a *distance measure* or a *data misfit function*) in order to measure how well these two images are similar under the transformation $\varphi(\mathbf{x})$. Then the general registration problem's aim is to minimise this measure in determining φ:

$$\text{Find } \varphi = (\varphi_1, \varphi_2)^\top \text{ such that } \mathcal{D}[T_\varphi, R, \varphi] = \mathcal{D}[\varphi] \text{ is minimal.} \tag{3}$$

We adopt the SSD or the least-squared function $\mathcal{D}$ defined by

$$\mathcal{D}[T_\varphi, R, \varphi] = \frac{1}{2}\int_\Omega (T(\varphi(\mathbf{x})) - R(\mathbf{x}))^2 \, d\mathbf{x} = \frac{1}{2}\|T_\varphi - R\|_{L_2}^2 = \mathcal{D}[\varphi] \tag{4}$$

as the optimisation object functional, where $\|\cdot\|_{L_2}$ denotes the L_2-norm.

2.1. Affine transformation.

Affine transformation is one of the most commonly used methods in registering two images (see [32], [15], [14], [23], [36], [38], [20], [29]). Although only linear, it models a combination of effects stemming from four simple transformations: translating, rotating, scaling and shearing. An affine transformation corrects some global distortions in the images to be registered. In this section, we first introduce the model and then discuss two numerical methods for solving it.

An affine registration model assumes that the above transformation φ is linear i.e.

$$\varphi(\mathbf{x}) = \varphi_{\mathbf{a}}(\mathbf{x}) = \begin{bmatrix} \varphi_{\mathbf{a}_1}(\mathbf{x}) \\ \varphi_{\mathbf{a}_2}(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ a_4 & a_5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} a_3 \\ a_6 \end{bmatrix} = \mathbf{A}\mathbf{x} + \mathbf{b}, \tag{5}$$

where $\mathbf{A} = \begin{bmatrix} a_1 & a_2 \\ a_4 & a_5 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} a_3 \\ a_6 \end{bmatrix}$ are the affine transformation matrix and the translation vector respectively, for all $\mathbf{x} \in \Omega$. Here for optimization purpose, the vector $\mathbf{a} = (a_1, a_2, a_3, a_4, a_5, a_6)^\top \in \mathbb{R}^6$ will be used shortly. Clearly the inverse transform is simply $\mathbf{x} = \mathbf{A}^{-1}(\varphi_{\mathbf{a}} - \mathbf{b})$ if $\mathbf{A}$ is invertible. Note that $\mathbf{A}$ can be decomposed into a product of a rotation, a scaling, a shear in x_1- (and/or x_2-)

direction or a combination of these simple transformations

$$\mathbf{A} = \begin{bmatrix} a_1 & a_2 \\ a_4 & a_5 \end{bmatrix} = \underbrace{\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}}_{\text{rotation}} \underbrace{\begin{bmatrix} s_{x_1} & 0 \\ 0 & s_{x_2} \end{bmatrix}}_{\text{scaling}} \underbrace{\begin{bmatrix} 1 & S_{x_1} \\ 0 & 1 \end{bmatrix}}_{\text{shear}} \tag{6}$$

where θ is the rotation angle, s_{x_1}, s_{x_2} are the scaling parameters, and S_{x_1} is the shear factor in x_1-direction. Clearly this kind of decomposition is not unique. It is clear that both a rigid-body transformation with $s_{x_1} = s_{x_2} = 1$ and $S_{x_1} = 0$ taking the form

$$\mathbf{A} = \begin{bmatrix} a_1 & a_2 \\ a_4 & a_5 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

and a similarity transformation with $0 < s_{x_1} = s_{x_2}$ and $S_{x_1} = 0$ taking the form

$$\mathbf{A} = \begin{bmatrix} a_1 & a_2 \\ a_4 & a_5 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} s_{x_1} & 0 \\ 0 & s_{x_2} \end{bmatrix}$$

are affine in special cases. From (4), the problem with such a φ is the affine image registration, formulated as follows

$$\min_{\mathbf{a}\in\mathbb{R}^6} \mathcal{D}[\mathbf{a}] \tag{7}$$

where $\mathcal{D}[\mathbf{a}] = \mathcal{D}[\varphi_{\mathbf{a}}] = \frac{1}{2}\|T(\varphi_{\mathbf{a}}) - R\|^2_{L_2} = \frac{1}{2}\|T(\mathbf{Ax}+\mathbf{b}) - R\|^2_{L_2}$.

Now consider how the discretized form of the minimisation problem (7) is solved, given discrete images $\mathbf{T}$ and $\mathbf{R}$ of T and R as $n_1 \times n_2$ arrays of image intensities. For ease of presentation, let $\mathbf{T}$ and $\mathbf{R}$, of dimension $N = n_1 n_2$, be pixel-wise ordered in a lexicographical order and denoted as follows

$$\mathbf{T} = (t_{1,1}, t_{2,1}, ..., t_{i_1,i_2}, ..., t_{n_1,n_2})^\top \quad \text{and} \quad \mathbf{R} = (r_{1,1}, r_{2,1}, ..., r_{i_1,i_2}, ...r_{n_1,n_2})^\top, \tag{8}$$

where $1 \le i_1 \le n_1$ and $1 \le i_2 \le n_2$. Each element in the grid vectors $\mathbf{T}$ and $\mathbf{R}$ represents a pixel's gray intensity between black (0) and white (1). Given an affine transformation $\varphi_{\mathbf{a}} = (\varphi_{\mathbf{a}_1}, \varphi_{\mathbf{a}_2})^\top$, the discrete form of the transformed template image F can be expressed as:

$$\begin{aligned} F(\mathbf{a}) = (&t_{a_1+a_2+a_3,\ a_4+a_5+a_6}, t_{2a_1+a_2+a_3,\ 2a_4+a_5+a_6}, ..., \\ &t_{a_1 i_1+a_2 i_2+a_3,\ a_4 i_1+a_5 i_2+a_6}, ..., t_{a_1 n_1+a_2 n_2+a_3,\ a_4 n_1+a_5 n_2+a_6})^\top, \end{aligned} \tag{9}$$

where $F : \mathbb{R}^6 \to \mathbb{R}^N$. Then problem (7) is equivalent to the following

$$\min_{\mathbf{a}\in\mathbb{R}^6} \mathcal{D}[\mathbf{a}] = \frac{1}{2}\|F(\mathbf{a}) - \mathbf{R}\|^2_{l_2} = \frac{1}{2}\|\mathbf{d}(\mathbf{a})\|^2_{l_2} = \frac{1}{2}\sum_{i=1}^{N} d_i(\mathbf{a})^2, \tag{10}$$

where we ignored the factor $h_1 h_2 = 1/(n_1 n_2) = 1/N$ due to discretization with the spatial mesh lengths h_1, h_2 but it will be used later in Section 5, $\mathbf{d}(\mathbf{a}) = F(\mathbf{a}) - \mathbf{R} \in \mathbb{R}^N$ is the so-called residual vector. The first order condition of (10) is

$$\mathbf{g}(\mathbf{a}) = \nabla_{\mathbf{a}} \mathcal{D}[\mathbf{a}] = \mathbf{J}^\top(\mathbf{a})(F(\mathbf{a}) - \mathbf{R}) = \mathbf{J}^\top \mathbf{d}(\mathbf{a}) = \mathbf{0}, \tag{11}$$

where $\mathbf{g}(\mathbf{a}) \in \mathbb{R}^6$ and the Jacobian matrix $\mathbf{J}$ given by

$$\mathbf{J}_{i,j} = \frac{\partial d_i}{\partial a_j} \qquad \text{for } 1 \le i \le N, 1 \le j \le 6. \tag{12}$$

Solving (11) for $\mathbf{a}$ is a non-linear problem and its solution requires an iterative approach. Let $\mathbf{a}^{(k)}$ be $\mathbf{a}$ at the kth iteration. Here, we must find a perturbation $\delta\mathbf{a}^{(k)}$ first and then update the solution vector by

$$\mathbf{a}^{(k+1)} = \mathbf{a}^{(k)} + \delta\mathbf{a}^{(k)}. \tag{13}$$

For a full Newton method, the perturbation $\delta\mathbf{a}^{(k)}$ is determined by solving

$$\mathbf{H}\left(\mathbf{a}^{(k)}\right)\delta\mathbf{a}^{(k)} = -\mathbf{g}\left(\mathbf{a}^{(k)}\right), \tag{14}$$

where the Hessian of $\mathcal{D}$ is denoted by

$$\mathbf{H}(\mathbf{a}) = \mathbf{J}^{\top}(\mathbf{a})\,\mathbf{J}(\mathbf{a}) + \sum_{i=1}^{N} d_i(\mathbf{a})\nabla^2 d_i(\mathbf{a}). \tag{15}$$

As pointed out by [23, p.79], this Newton method may be not suitable in registering two images for practical applications because computing higher order derivatives is time consuming and numerically unstable. In order to improve on the Newton method, we can take advantages of the particular structure of $\mathbf{H}$ to design a Newton variant to compute $\delta\mathbf{a}^{(k)}$.

2.2. The Gauss-Newton method. Note that the Hessian matrix is precisely $\mathbf{H}(\mathbf{a}) = \mathbf{J}^{\top}(\mathbf{a})\,\mathbf{J}(\mathbf{a})$ if $d_i = 0$ for all i (i.e. the residuals are zero at the solution $\mathbf{a}^*$) or if $\nabla^2 d_i(\mathbf{a}) = 0$ when d_i is a linear function of $\mathbf{a}$. This suggests that in other cases the Hessian matrix may also be approximated by this formula [30]. The resulting approximation leads to the Gauss-Newton (GN) method, defined by

$$\widetilde{\mathbf{H}}\left(\mathbf{a}^{(k)}\right)\delta\mathbf{a}^{(k)} = -\mathbf{g}\left(\mathbf{a}^{(k)}\right), \tag{16}$$

where one uses the matrix $\widetilde{\mathbf{H}}\left(\mathbf{a}^{(k)}\right) = \mathbf{J}^{\top}\left(\mathbf{a}^{(k)}\right)\mathbf{J}\left(\mathbf{a}^{(k)}\right)$ to approximate $\mathbf{H}\left(\mathbf{a}^{(k)}\right)$.

The above GN method requires damping to ensure convergence, because we may not be able to provide a good initial solution, close to a minimum of $\mathcal{D}$. The damped GN method can be generated by

$$\mathbf{a}^{(k+1)} = \mathbf{a}^{(k)} + \alpha^{(k)}\delta\mathbf{a}^{(k)} \tag{17}$$

where the positive scalar $\alpha^{(k)}$ is the so-called *line-search* parameter used to ensure that a GN step adequately reduces $\mathcal{D}$ and to rule out an unacceptable short step. More precisely, $\alpha^{(k)}$ is determined by

$$\alpha^{(k)} = \arg\min_{\alpha} \mathcal{D}\left[\mathbf{a}^{(k)} + \alpha\delta\mathbf{a}^{(k)}\right].$$

Solving this line-search problem is by a *backtracking* algorithm which begins with $\alpha^{(k)} = 1$, and then, if $\mathbf{a}^{(k)} + \delta\mathbf{a}^{(k)}$ is not acceptable, reduces $\alpha^{(k)}$ until an acceptable $\mathbf{a}^{(k)} + \alpha^{(k)}\delta\mathbf{a}^{(k)}$ is found. The acceptability is decided by the so-called Wolfe or Armijo-Goldstein conditions safeguarding upper and lower bounds; see Dennis and Schnabel [3], Fletcher [6], Kelly [16], Nocedal and Wright [25], and Yuan and Sun [37].

2.3. The Levenberg-Marquardt method. The GN method (16) assumes that $\widetilde{\mathbf{H}}\left(\mathbf{a}^{(k)}\right)$ is well-conditioned or at least non-singular. To remove this assumption, an improved formulation by adding a positive multiple of the identity matrix $\mathbf{I}$, $\widetilde{\mathbf{H}}\left(\mathbf{a}^{(k)}\right) = \left[\mathbf{J}^{\top}\left(\mathbf{a}^{(k)}\right)\mathbf{J}\left(\mathbf{a}^{(k)}\right) + \mu^{(k)}\mathbf{I}\right]$, is the Levenberg-Marquardt (LM) method:

$$\left[\mathbf{J}^{\top}\left(\mathbf{a}^{(k)}\right)\mathbf{J}\left(\mathbf{a}^{(k)}\right) + \mu^{(k)}\mathbf{I}\right]\delta\mathbf{a}^{(k)} = -\mathbf{g}\left(\mathbf{a}^{(k)}\right). \tag{18}$$

At iteration k, the positive LM parameter $\mu^{(k)}$ is adjusted to guarantee that the search direction $\delta\mathbf{a}^{(k)}$ in (18) is a descent direction. Then we get a steepest descent direction for large $\mu^{(k)}$, when the current iterate is far from the solution. On the other hand, this descent direction is approximately a GN search direction for small $\mu^{(k)}$, when the iterates get close enough to the solution. Using a frame work of trust region strategies, the LM parameter $\mu^{(k)}$ is determined in such a way that

$$\left\|\delta\mathbf{a}^{(k)}\right\|_{l_2}^2 = \left\|\left[\mathbf{J}^\top\left(\mathbf{a}^{(k)}\right)\mathbf{J}\left(\mathbf{a}^{(k)}\right)+\mu^{(k)}\mathbf{I}\right]^{-1}\left[-\mathbf{g}\left(\mathbf{a}^{(k)}\right)\right]\right\|_{l_2}^2 \leq \eta^{(k)} \tag{19}$$

where $\eta^{(k)} \geq 0$ is a prescribed trust region radius. A new LM step is then generated by $\mathbf{a}^{(k+1)} = \mathbf{a}^{(k)} + \delta\mathbf{a}^{(k)}$. As remarked by Hanke [10], Henn [11], Doicu *et al.* [4, 5], this numerical scheme is related to Tikhonov regularization (see Section 5) and is sometimes called the regularizing Levenberg-Marquardt method as shown in (43).

2.4. Some registration results using the GN and LM methods. In this section, some registration results using the GN method (Section 2.2) and the LM method (Section 2.3) are presented, to illustrate the non-robustness of both GN and LM methods.

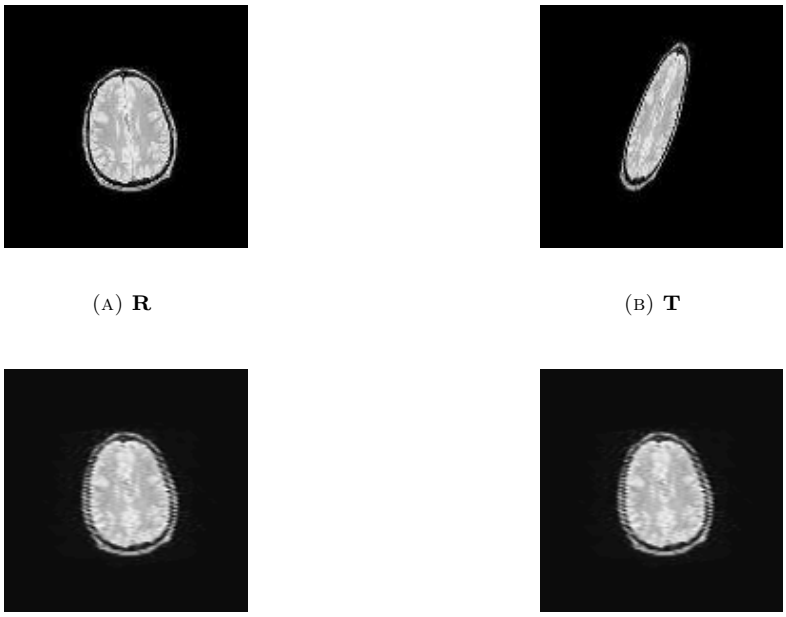
We shall give two examples, with the first one to show that both methods are capable of correctly registering 2 images and the second one to show that both methods can fail to converge to an acceptable solution i.e. fail to register 2 images (in particular our examples will differ in outliers). In both examples, the images are of size 128×128 and for both GN and LM, we use the termination criterion $\|\delta\mathbf{a}\|_{l_2} \leq \epsilon = 10^{-6}$ within the maximum of iterations IMAX= 300. The bilinear interpolation technique was applied in all examples for computing the transformed template image $F(\mathbf{a}) = \mathbf{T}_{\varphi_\mathbf{a}}$. We shall measure the relative residual as the error indicator: $error = \|\mathbf{g}(\mathbf{a}^{(\mathbf{k})})\|_{l_2}$.

Example 1 (A successful case)**.** *We consider the registration problem for a pair of MR images of a human head*[1]*, with the reference image* $\mathbf{R}$ *and the template image* $\mathbf{T}$ *respectively in Figure 1 (a)–(b). Using the initial guess* $\mathbf{a}^{(0)} = (1,0,0,0,1,0)^\top$ *(i.e. we start with* $\varphi_0(\mathbf{x}) = \mathbf{x}$*), both the GN and LM methods can successfully register this example as shown respectively in Figure 1 (c)–(d).*

Example 2 (An unsuccessful case)**.** *Here we consider another pair of MR images (similar to Example 1), as shown in Figure 2 (a)–(b), where* $\mathbf{T}$ *contains tumor like circles. As in Example 1, the initial guess solution* $\mathbf{a}^{(0)} = (1,0,0,0,1,0)^\top$ *is used. It turns out that both GN and LM methods get stuck (at a local minimum of* $\mathcal{D}$*) and fail to obtain correctly converged solutions, as shown in Figure 2 (c)–(d). Here we are certain about reaching a local minimum because the residual error is small, and about the registration failure because we can observe the large visual difference between* $F(\mathbf{a})$ *and* $\mathbf{R}$ *(i.e. the matching error is not the smallest possible).*

Based on Examples 1 and 2 and other tests, we confirm that (as is known) both methods are not robust enough as their convergence strongly depends on initial guess solutions. We shall shortly discuss various ways of finding good initial guess solutions.

[1]Source: http://www.cis.rit.edu/class/schp730/lect/lect-1.htm



(A) $\mathbf{R}$ (B) $\mathbf{T}$

(C) $F_{\mathrm{GN}}(\mathbf{a}^{(\mathbf{98})})$ *(error=2.77*$\times 10^{-7}$*)*

(D) $F_{\mathrm{LM}}(\mathbf{a}^{(\mathbf{107})})$ *(error=*3.38×10^{-7}*)*

FIGURE 1. Example 1: Successful registration results of the MR images of a human head. The first row shows the reference image $\mathbf{R}$ (a), the template image $\mathbf{T}$ (b). The second row presents the registered images $F_{\mathrm{GN}}(\mathbf{a}^{(\mathbf{98})})$ (c) and $F_{\mathrm{LM}}(\mathbf{a}^{(\mathbf{107})})$ (d) obtained from using the GN and LM methods, respectively.

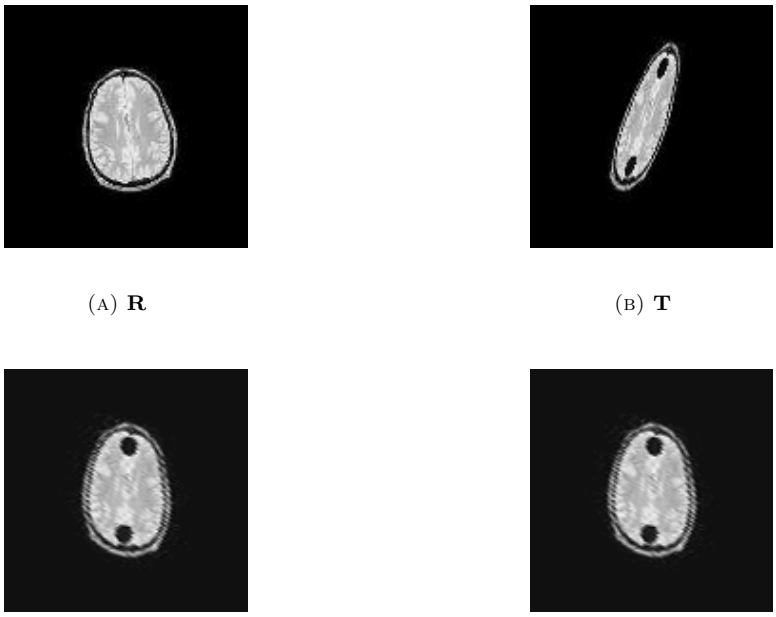
3. Deformable registration

Having discussed a parametric registration model, we now give a brief review of a non-parametric model – the variational diffusion model for deformable registration [23]. We shall show that, although the nonlinear multigrid method [33] is effective in solving the model, an affine pre-registration step can further speed up the solution. Hence it is of interest to look for reliable affine methods. We first review the general Tikhonov regularization idea [1, 23, 26, 34].

3.1. Variational approach. As an inverse problem, the general registration problem (3) denoted by $\min_{\varphi} \mathcal{D}[\varphi]$ is ill-posed and can be converted to a well-posed problem by Tikhonov regularization leading to

$$\min_{\varphi} \mathcal{J}_{\beta}[\varphi] = \mathcal{D}[\varphi] + \beta \mathcal{S}[\mathbf{x} - \varphi] \tag{20}$$

where the positive regularizer $\mathcal{S}$ may be chosen differently [23], and $\beta > 0$ is the *regularization parameter*, which controls the fitting of the registered image, as measured by the term $\mathcal{D}[\varphi]$, and the regularity of the solution, as measured by the term $\mathcal{S}[\mathbf{x} - \varphi]$.



(A) $\mathbf{R}$ (B) $\mathbf{T}$

(C) $F_{\mathrm{GN}}(\mathbf{a^{(65)}})$ *(error=1.88*$\times 10^{-7}$*)*

(D) $F_{\mathrm{LM}}(\mathbf{a^{(103)}})$ *(error=2.47*$\times 10^{-7}$*)*

FIGURE 2. Example 2: Unsuccessful registration results of the MR images of a human head. The first row shows the reference image $\mathbf{R}$ (a), the template image $\mathbf{T}$ (b). The second row presents the registered images $F_{\mathrm{GN}}(\mathbf{a^{(65)}})$ (c) and $F_{\mathrm{LM}}(\mathbf{a^{(103)}})$ (d) obtained from using the GN and LM methods, respectively.

To have a consistent notation with the literature, define the new *deformation* variable $\mathbf{u}(\mathbf{x}) = \mathbf{x} - \varphi(\mathbf{x})$. Then the variational problem (20) becomes

$$\min_{\mathbf{u}} \mathcal{J}_\beta[\mathbf{u}] = \mathcal{D}[\mathbf{u}] + \beta \mathcal{S}[\mathbf{u}], \tag{21}$$

where $\mathcal{D}[\mathbf{u}] = \frac{1}{2}\int_\Omega \left(T(\mathbf{x} - \mathbf{u}(\mathbf{x})) - R(\mathbf{x})\right)^2 d\mathbf{x}$.

3.2. Diffusion registration. The diffusion registration by Fischer and Modersitzki [7] chose the following *diffusive regularizer*

$$\mathcal{S}[\mathbf{u}] = \frac{1}{2}\int_\Omega \Big(|\nabla u_1|^2 + |\nabla u_2|^2 \Big) d\mathbf{x}, \tag{22}$$

subject to Neumann boundary conditions, i.e.,

$$\frac{\partial u_\ell}{\partial \vec{n}} = 0 \quad \text{for } \mathbf{x} \in \partial\Omega \text{ and } \ell = 1, 2. \tag{23}$$

Here, $\vec{n}$ denotes the unit outer normal vector on $\partial\Omega$. The Euler-Lagrange equation for the variational problem (21) is the following

$$\beta \Delta \mathbf{u}(\mathbf{x}) + \Big(T(\mathbf{x} - \mathbf{u}(\mathbf{x})) - R(\mathbf{x})\Big) \cdot \nabla T(\mathbf{x} - \mathbf{u}(\mathbf{x})) = 0, \quad \mathbf{x} \in \Omega, \tag{24}$$

where the Gâteaux-derivatives of $\mathcal{S}$ are used and Δ denotes the Laplace operator with $\Delta\mathbf{u}(\mathbf{x}) = (\Delta u_1, \Delta u_2)$. Note that (24) denotes a system of two non-linear PDEs.

3.3. Numerical treatment and results. In [7], the cell-centered finite difference scheme is recommended to discretize the parabolic version of (24) i.e.

$$\frac{\partial \mathbf{u}}{\partial t} = \beta \Delta \mathbf{u}(\mathbf{x}) + \Big(T(\mathbf{x} - \mathbf{u}(\mathbf{x})) - R(\mathbf{x})\Big) \cdot \nabla T(\mathbf{x} - \mathbf{u}(\mathbf{x}))$$

and solve the discrete system by the so-called additive operator splitting (AOS) method [19, 35] which is a semi-implicit time marching method.

Below we shall apply the finite difference method to (24) directly and report on some results from adopting a full approximation scheme multigrid (with full multigrid initialization) method, denoted by FMG-FAS, as in [33]. The basic steps are briefly summarized as follows: (i) Convert the original fine grid problem to a hierarchy of coarser levels with standard coarsening. The linearized Gauss-Seidel smoother is employed for (24), while on the coarsest level the AOS-scheme of [7] is used. We take the number of pre- and post-smoothing to be 3, and the number of outer iterations to be 2. (ii) Use the standard bi-linear interpolation and restriction operators.

Example 3. *We consider the deformable registration problem of the X-Ray images of a human hand*[2]*. Figures 3 (a)−(b) show the reference* $\mathbf{R}$ *and template* $\mathbf{T}$ *images. Clearly one can tell that the two images are not related by affine transforms. However we use an affine transform to provide a good initial guess which we denote by* $\mathbf{T}^{\mathbf{lin}}_{GN}$ *in Figure 3 (c), obtained from the affine method as in Section 2.2.*

Then the registered images $F(u)$ *obtained from (24) by the FMG-FAS method with and without the affine pre-registration step are shown, respectively, in Figures 3 (e)−(f). The latter method (without using the affine pre-registration step) is not only much slower than the former with the affine step (only* 2 *FAS cycles), but also it failed to register properly (Figure 3 (f)). Here we remark that without the affine pre-registration step, essentially, it is the FMG that struggles on the coarsest grid.*

Through the above example, we see that a deformable registration approach can benefit from an affine pre-registration step whose convergence is of course of importance.

4. Techniques to improve affine registration methods

The convergence of both GN and LM methods depends on suitable initial guesses, as shown in Section 2.4. Below we shall first survey several methods that can provide a better initial guess than the simple $\mathbf{a}^{(0)} = (1, 0, 0, 0, 1, 0)^\top$ for the GN affine model (16), and then test these methods for their usefulness to both GN and LM methods.

4.1. Method 1 − **Approximation based on image centers.** Each of the two input images has a center location, defined by the pixel gray levels (or features dependent). If the two centers are quite different, re-positioning the center will give the affine registration problem a good initial guess for translation i.e. the vector given by $\mathbf{a}^{(0)} = (1, 0, a_3^{(0)}, 0, 1, a_6^{(0)})^\top$, where $a_3^{(0)}$ and $a_6^{(0)}$ denote the re-positioning information.

[2]Source: http://www.math.mu-luebeck.de/safir/

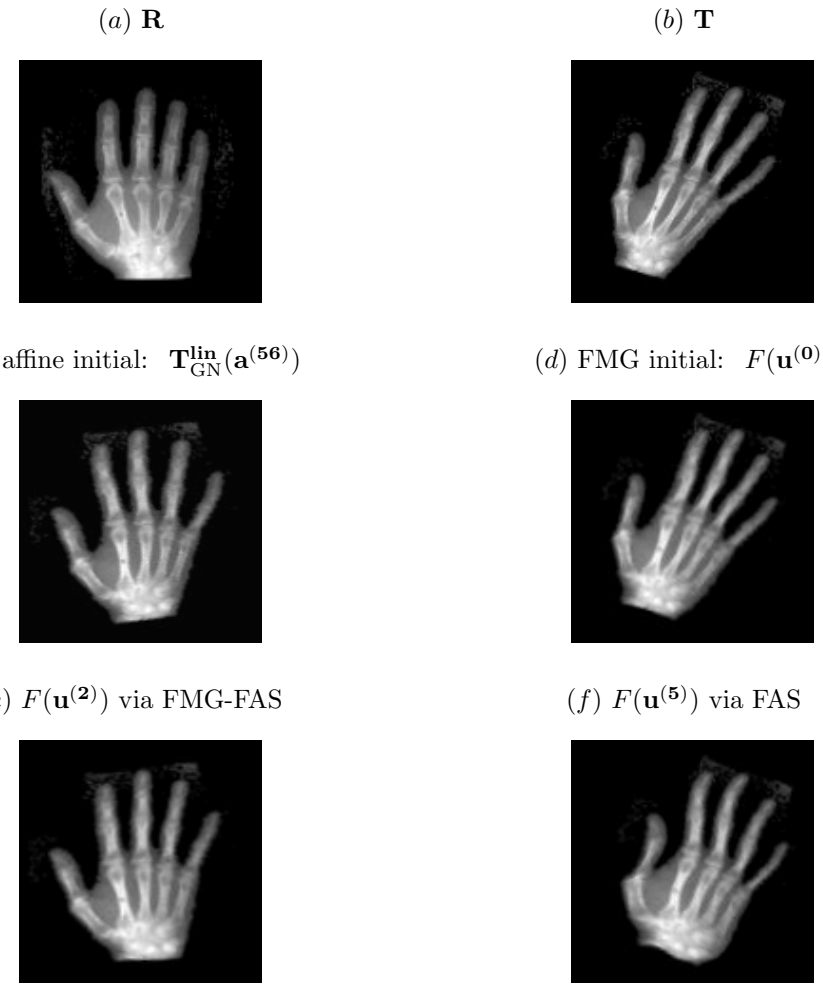


FIGURE 3. Example 3: Deformable registration results of the X-Ray images of a human hand, showing the importance of a pre-registration step. Left: (a) Reference $\mathbf{R}$, (c) the linearly registered template (initial template) using the GN method $\mathbf{T}^{\mathbf{lin}}_{\mathrm{GN}}(\mathbf{a}^{(\mathbf{56})})$, and (e) the registered image $F(\mathbf{u}^{(\mathbf{2})})$ by FMG-FAS with (c). Right: (b) Template $\mathbf{T}$, (d) the initial image $F(\mathbf{u}^{(\mathbf{0})})$ after FMG step, and (f) the (failed) registered image $F(\mathbf{u}^{(\mathbf{5})})$ with (d).

This method can be summarized as follows.

(i) Estimate the centers c^T, c^R of the two input images $T,\ R$ respectively:

$$\text{(25)}\quad \begin{cases} c^T = \begin{bmatrix} c_1^T \\ c_2^T \end{bmatrix} = \frac{\int_\Omega \mathbf{x} T(\mathbf{x}) d\mathbf{x}}{\int_\Omega T(\mathbf{x}) d\mathbf{x}} = \frac{(\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} t_{i,j}\, i,\ \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} t_{i,j}\, j)}{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} t_{i,j}}, \\ c^R = \begin{bmatrix} c_1^R \\ c_2^R \end{bmatrix} = \frac{\int_\Omega \mathbf{x} R(\mathbf{x}) d\mathbf{x}}{\int_\Omega R(\mathbf{x}) d\mathbf{x}} = \frac{(\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} r_{i,j}\, i,\ \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} r_{i,j}\, j)}{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} r_{i,j}} \equiv \mathbb{E}_R\left[\mathbf{x}\right]. \end{cases}$$

(ii) From the center differences, set $a_3^{(0)} = c_1^R - c_1^T,\ a_6^{(0)} = c_2^R - c_2^T$.

4.2. Method 2 – Approximation based on the rigid-body model . The next idea of providing a good initial guess for (5) is to reduce the number of parameters: assume there exists a rigid transform between T and R. Then we have an parametric model $\varphi(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ with only 3 parameters (see (6)):

$$\mathbf{A} = \begin{bmatrix} a_1 & a_2 \\ a_4 & a_5 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} a_3 \\ a_6 \end{bmatrix}. \tag{26}$$

Here we could use the above Method 1 to initialize $\mathbf{b}$ while setting $\theta^{(0)} = 0$. Once this model is solved, we shall supply the coefficients to $\mathbf{a}^{(0)}$ for the affine model.

4.3. Method 3 – Approximation based on principal axes transformation . The *principal axes transformation* (PAT) method was introduced to image processing by Hu since 1962 (see [23, 29] and references therein). It is an approxmiate registration approach using statistical features, the image center and an eigendecomposition of the covariance matrix, derived from the input images. Define the 2×2 covariance matrix of an image I by

$$Cov_I = \mathbb{E}_I \left[\left(\mathbf{x} - c^I\right) \left(\mathbf{x} - c^I\right)^\top \right], \tag{27}$$

where c^I is the image center defined by (25). Since matrix Cov_I is real, symmetric, and positive semi-definite, it permits an eigenvalue decomposition [23]

$$Cov_I = D(\rho_I)\, \Sigma_I^2 D(-\rho_I), \quad D(\rho_I) = \begin{bmatrix} \cos\rho_I & -\sin\rho_I \\ \sin\rho_I & \cos\rho_I \end{bmatrix}, \quad \Sigma_I = \begin{bmatrix} \sigma_{I,1} & 0 \\ 0 & \sigma_{I,2} \end{bmatrix} \tag{28}$$

where $D(\rho_I)$ denotes a rotation matrix, Σ_I is a scaling matrix, and $\sigma_{I,1}$ and $\sigma_{I,2}$ are standard deviations.

Now for images R, T, let c^R and c^T be the centers from (25). Then the following will be the approximate coefficients for an affine transform

$$\mathbf{A} = D(\rho_T)\, \Sigma_T \Sigma_R^{-1} D(\rho_R)^\top, \quad \mathbf{b} = c^T - \mathbf{A}c^R. \tag{29}$$

Finally the coefficients from (29) will be used to initialize $\mathbf{a}^{(0)}$ for the affine model.

4.4. Method 4 – Multi-resolution approach. Multi-resolution strategy is commonly used to provide reliable initial guesses for registration algorithms [15, 18, 21, 28, 31, 32]. The idea is to register the coarse resolution (low) images first and then interpolate the coarse solutions level by level to the finest resolution (high). The basic idea is essentially the same as a full multigrid method as in [33] and done in Section 3.3.

Suppose that we operate with L levels in total (using standard coarsening [33]), with $\ell = 1$ the coarsest level and $\ell = L$ the finest level. Here the size of the coarsest level 1 is chosen as 32×32 or 64×64, and the bi-linear interpolation is used. Although the full weighting operator [33] may be used for restriction, the usual practice is to use a Gaussian-like kernel typically consisted of a 5×5 template of weights as follows. Take the reference image $\mathbf{R} = \mathbf{R}_L$ as example. Define a coarsening operation from $\mathbf{R}_\ell$ to $\mathbf{R}_{\ell-1}$, i.e. $\mathbf{R}_{\ell-1} = coarsen(\mathbf{R}_\ell)$, by

$$R_{\ell-1}(i,j) = \sum_{k_1=-2}^{2} \sum_{k_2=-2}^{2} w(k_1) w(k_2) R_\ell(2i + k_1, 2j + k_2),$$

where $w(0) = 2/5$, $w(\pm 1) = 1/4$, and $w(\pm 2) = 1/4 - w(0)/2$. On level ℓ a standard nonlinear least squares method (either GN or LM) is used to compute the affine transformation up to some tolerance (e.g. $tol = 10^{-2}$), which is denoted by $\mathbf{a}_\ell \leftarrow$

Solver_Step$(\mathbf{T}_\ell, \mathbf{R}_\ell, \mathbf{a}_\ell)$. Then the whole procedure of Method 4 may be denoted by $\mathbf{a}_L \leftarrow multiresolution(\mathbf{T}_L, \mathbf{R}_L, \mathbf{a}_L, L)$ with a recursion step summarized below:

Algorithm 1 (Multi-resolution approach)**.**
Implement $\mathbf{a}_\ell \leftarrow multiresolution(\mathbf{T}_\ell, \mathbf{R}_\ell, \mathbf{a}_\ell, \ell)$ *as follows:*

- If $\ell = 1$
  - Set $\mathbf{a}_\ell = (1, 0, 0, 0, 1, 0)^\top$ or use Methods 1-3 to work out an initial $\mathbf{a}_\ell$,
  - $\mathbf{a}_\ell \leftarrow$ *Solver_Step*$(\mathbf{T}_\ell, \mathbf{R}_\ell, \mathbf{a}_\ell)$.
- Else
  - $\mathbf{T}_{\ell-1} = coarsen\,(\mathbf{T}_\ell)\,, \quad \mathbf{R}_{\ell-1} = coarsen\,(\mathbf{R}_\ell)\,.$
  - $\mathbf{a}_{\ell-1} \leftarrow multiresolution(\mathbf{T}_{\ell-1}, \mathbf{R}_{\ell-1}, \mathbf{a}_{\ell-1}, \ell - 1)$
  - $\mathbf{a}_\ell \leftarrow interpolate(\mathbf{a}_{\ell-1})$ as follows:
    $a_{i_\ell} = a_{i_{\ell-1}}$ for $i = 1, 2, 4, 5$ (the elements of the affine transformation matrix) and $a_{i_\ell} = 2a_{i_{\ell-1}}$ for $i = 3, 6$ (the elements of the translation matrix).
  - $\mathbf{a}_\ell \leftarrow$ *Solver_Step*$(\mathbf{T}_\ell, \mathbf{R}_\ell, \mathbf{a}_\ell)$.

4.5. Applications of Methods 1-4 to GN and LM methods. To illustrate the performance of Methods 1-4, we give two successful examples: firstly re-solve Example 2 and secondly consider a new Example 4.

Recall from Figure 2 that both GN and LM methods failed to converge to the desirable solution for Example 2 with a simple initial guess. Now with Methods 1-4 to provide initial guesses, both GN and LM methods work successfully – we show the registered results from GN in Figure 4 (while the LM results are virtually identical).

Example 4. *Here we consider the deformable registration problem for a pair of MR images of a human head, with Figure 5 (a)−(b) showing the reference image* $\mathbf{R}$ *and the template image* $\mathbf{T}$ *in size* 128×128.

As this is a deformable (not affine) problem, we can only use Methods 1-4 *to provide an initial guess solution for the affine model, whose solution is then used for the diffusion registration method of Section 3.3. Figure 5 (c)−(d) shows the results of affine GN and LM methods with Method* 4 *providing the initial guess, with the GN taking only* 5 *iterations and the LM taking* 11 *iterations on the finest resolution. Further, using Figure 5 (c)−(d) as initial guesses, the diffusion registration method of Section 3.3 with* $\beta = 0.058$ *gives the respectively registered images as depicted in Figure 5 (e)−(f).*

Of the four methods, Method 4 is believed to be the best because Methods 1-2 are not as general as 3-4, and Method 3 is unable to resolve shear components [29]. However, even Method 4 cannot provide good initial guesses for some examples, as shown later in Example 6. Although one can think about designing better ways than Methods 1-4 to provide more reliable and robust initial guesses for the affine model, our idea below is to propose a modified affine model that is less demanding than the standard model for initial guesses.

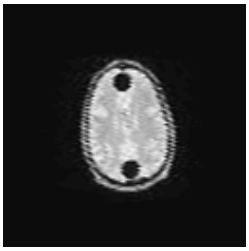
5. A regularized affine registration model

We propose to regularise the minimizing functional. Although the regularization idea is widely known for non-parametric models (as in Section 3.1), it is not usually applied to parametric registration problems.

Motivated by (20), we solve, instead of (3), the following minimisation problem:

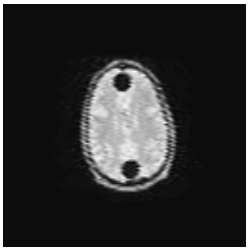
$$\min_{\varphi_{\mathbf{a}}} \mathcal{J}_\beta\,[\varphi_{\mathbf{a}}] = \mathcal{D}\,[\varphi_{\mathbf{a}}] + \bar{\beta}\mathcal{S}\,[\mathbf{x} - \varphi_{\mathbf{a}}]\,, \tag{30}$$

(a) Method 1 $F_{\mathrm{GN}}(\mathbf{a}^{(\mathbf{82})})$



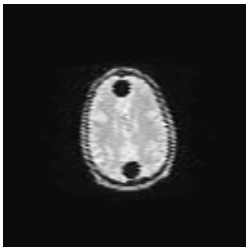
($error = 2.34 \times 10^{-6}$)

(b) Method 2 $F_{\mathrm{GN}}(\mathbf{a}^{(\mathbf{78})})$



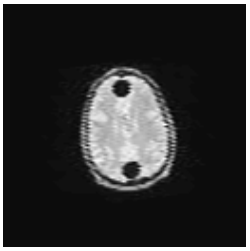
($error = 1.79 \times 10^{-6}$)

(c) Method 3 $F_{\mathrm{GN}}(\mathbf{a}^{(\mathbf{45})})$



($error = 6.93 \times 10^{-7}$)

(d) Method 4 $F_{\mathrm{GN}}(\mathbf{a}^{(\mathbf{38})})$



($error = 7.01 \times 10^{-7}$)

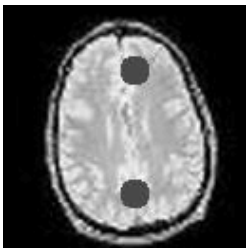
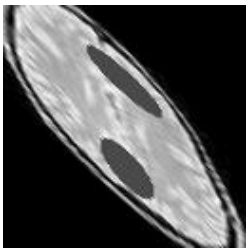
FIGURE 4. Example 2 re-solved: Correct registration results using the GN method with Methods 1-4 providing initial guess solutions respectively for (a), (b), (c) and (d).

where the regularizer $\mathcal{S}$ for affine image registration is proposed to take the form (31)

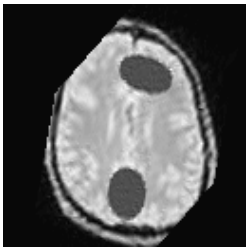
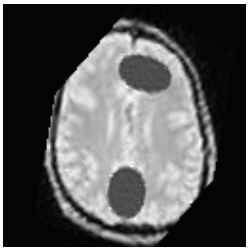
$$\mathcal{S}\left[\mathbf{x}-\varphi_{\mathbf{a}}\right]=\begin{cases}\mathcal{S}_1=\dfrac{1}{2}\displaystyle\sum_{i=1}^{2}\left\|x_i-\varphi_{\mathbf{a}_i}(\mathbf{x})\right\|_{H^1_{semi}}^2=\dfrac{1}{2}\sum_{i=1}^{2}\int_\Omega\Big(|\nabla_{x_i}(x_i-\varphi_{\mathbf{a}_i}(\mathbf{x}))|_{L_2}^2\Big)d\Omega,\\ \mathcal{S}_2=\dfrac{1}{2}\displaystyle\sum_{i=1}^{2}\left\|x_i-\varphi_{\mathbf{a}_i}(\mathbf{x})\right\|_{L_2}^2=\dfrac{1}{2}\sum_{i=1}^{2}\int_\Omega\Big((x_i-\varphi_{\mathbf{a}_i}(\mathbf{x}))^2\Big)d\Omega,\\ \mathcal{S}_3=\mathcal{S}_1\left[\mathbf{x}-\varphi_{\mathbf{a}}\right]+\mathcal{S}_2\left[\mathbf{x}-\varphi_{\mathbf{a}}\right],\\ \mathcal{S}_4=\dfrac{1}{2}\left\|\mathbf{a}\right\|_{l_2}^2.\end{cases}$$

Here the regularizers $\mathcal{S}_1$, $\mathcal{S}_2$, and $\mathcal{S}_3$ are motivated by regularization of non-affine models, differing only in norms for functions, which are respectively the Sobolev semi-norm H^1_{semi}, the L_2-norm, and the H^1 norm. $\mathcal{S}_4$ is a simple option, using the l_2-norm. Clearly the new regularized affine registration (RAR) model (30) reduces to be the classical one (3) when $\bar{\beta}=0$. As in Section 3.1, the regularization parameter $\bar{\beta}$ balances the influence of $\mathcal{D}$ and $\mathcal{S}$. We shall discuss how to choose it shortly in the next section. For affine problems where the true solutions require large translations, one may argue that such regularization might restrict solutions from reaching true solutions. Fortunately our tests will show that this is not the case.

(*a*) Reference $\mathbf{R}$ (*b*) Template $\mathbf{T}$

(*c*) GN with Method 4 (*d*) LM with Method 4

(*e*) FAS $F(\mathbf{u}^{(\mathbf{6})})$ + GN (*f*) FAS $F(\mathbf{u}^{(\mathbf{3})})$ + LM

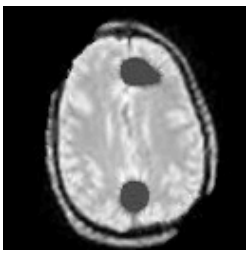
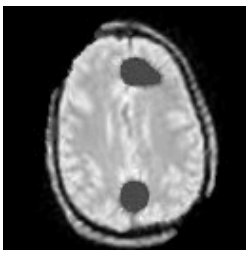



FIGURE 5. Example 4: Correct registration results of the MR images of a human head (deformable model of Section 3.2 with initial solutions provided by Method 4) as in row 1. The second row displays the helpful pre-registration images obtained from (*c*) the GN $\mathbf{T}_{\mathrm{GN}}$ and (*d*) the LM method $\mathbf{T}_{\mathrm{LM}}$. The last row (e)-(f) shows the deformable model (via FAS) registered images starting with (c)-(d) respectively.

We now express the proposed regularizers in an analytical form in the terms of the six-parameter vector $\mathbf{a} = (a_1, a_2, a_3, a_4, a_5, a_6)^\top$ as follows:

$$\mathcal{S}_1[\mathbf{a}] = \frac{1}{2}\left((1-a_1)^2 + a_2^2 + a_4^2 + (1-a_5)^2\right), \tag{32}$$

$$\begin{aligned}\mathcal{S}_2[\mathbf{a}] = \frac{1}{2}\Bigg(&\frac{a_1^2}{3} + \frac{a_2^2}{3} + a_3^2 + \frac{a_4^2}{3} + \frac{a_5^2}{3} + a_6^2 \\ &+ \frac{1}{2}(a_1a_2 + a_4a_5) + a_1a_3 + a_2a_3 + a_4a_6 + a_5a_6 \\ &- \frac{2}{3}(a_1 + a_5) - \frac{1}{2}(a_2 + a_4) - (a_3 + a_6) + \frac{2}{3}\Bigg),\end{aligned} \tag{33}$$

$$\begin{aligned}\mathcal{S}_3\left[\mathbf{a}\right] = \frac{1}{2}\Bigg(&\frac{4}{3}a_1^2+\frac{4}{3}a_2^2+a_3^2+\frac{4}{3}a_4^2+\frac{4}{3}a_5^2+a_6^2\\&+\frac{1}{2}\left(a_1a_2+a_4a_5\right)+a_1a_3+a_2a_3+a_4a_6+a_5a_6\\&-\frac{8}{3}\left(a_1+a_5\right)-\frac{1}{2}\left(a_2+a_4\right)-\left(a_3+a_6\right)+\frac{8}{3}\Bigg),\end{aligned} \tag{34}$$

$$\mathcal{S}_4\left[\mathbf{a}\right] = \frac{1}{2}\left(a_1^2+a_2^2+a_3^2+a_4^2+a_5^2+a_6^2\right). \tag{35}$$

Further apply the GN approach to solve the discrete minimisation problem:

$$\min_{\mathbf{a}\in\mathbb{R}^6}\mathcal{J}_\beta\left[\mathbf{a}\right] = \mathcal{D}\left[\mathbf{a}\right]+\bar{\beta}N\mathcal{S}\left[\mathbf{a}\right] = \mathcal{D}\left[\mathbf{a}\right]+\beta\mathcal{S}\left[\mathbf{a}\right], \tag{36}$$

where the factor $N = n^2$ for a square image $n\times n$ is now needed for multi-resolution setting with $\beta = \bar{\beta}N$, since the discrete fitting term $\mathcal{D}$ as in (10) does not contain step-lengths information. The GN perturbation $\delta\mathbf{a}^{(k)}$ for (36) is then given by

$$\widetilde{\mathbf{H}}_{\mathcal{J}_\beta}\left(\mathbf{a}^{(k)}\right)\delta\mathbf{a}^{(k)} = -\mathbf{g}_{\mathcal{J}_\beta}\left(\mathbf{a}^{(k)}\right) \tag{37}$$

where

$$\widetilde{\mathbf{H}}_{\mathcal{J}_\beta}\left(\mathbf{a}^{(k)}\right) = \mathbf{J}^\top\left(\mathbf{a}^{(k)}\right)\mathbf{J}\left(\mathbf{a}^{(k)}\right)+\beta\mathbf{H}_{\mathcal{S}}\left(\mathbf{a}^{(k)}\right) \tag{38}$$

and

$$\mathbf{g}_{\mathcal{J}_\beta}\left(\mathbf{a}^{(k)}\right) = \mathbf{g}\left(\mathbf{a}^{(k)}\right)+\beta\nabla_{\mathbf{a}}\mathcal{S}[\mathbf{a}^{(k)}] \tag{39}$$

are the approximated Hessian and the gradient of $\mathcal{J}_\beta$ at $\mathbf{a}^{(k)}$, and $\nabla_{\mathbf{a}}\mathcal{S}[\mathbf{a}^{(k)}]$ and $\mathbf{H}_{\mathcal{S}}\left(\mathbf{a}^{(k)}\right)$ are respectively the gradient and the Hessian of $\mathcal{S}$ at $\mathbf{a}^{(k)}$. Note that for $\mathcal{S} = \mathcal{S}_1$ we may approximate $\mathbf{H}_{\mathcal{S}_1}\left(\mathbf{a}^{(k)}\right)$ by $\mathbf{I}$ because it helps $\widetilde{\mathbf{H}}_{\mathcal{J}_\beta}\left(\mathbf{a}^{(k)}\right)$ to be a symmetric positive definite matrix. As before, once we have the GN update $\delta\mathbf{a}^{(k)}$, we can also apply the line-search idea: $\mathbf{a}^{(k+1)} = \mathbf{a}^{(k)}+\alpha^{(k)}\delta\mathbf{a}^{(k)}$.

A connection between our RAR method and the LM method from Section 2.3 can be explained as follows. Consider the regularizer $\mathcal{S}_4$ with a fixed β. Our RAR method defines the perturbation given by

$$\left[\mathbf{J}^\top\left(\mathbf{a}^{(k)}\right)\mathbf{J}\left(\mathbf{a}^{(k)}\right)+\beta\mathbf{I}\right]\delta\mathbf{a}^{(k)} = -\left[\mathbf{g}\left(\mathbf{a}^{(k)}\right)+\beta\mathbf{I}\mathbf{a}^{(k)}\right], \tag{40}$$

which is a solution of the following minimisation problem:

$$\min_{\delta\mathbf{a}\in\mathbb{R}^6}\frac{1}{2}\left\|F(\mathbf{a}^{(k)})+\mathbf{J}\left(\mathbf{a}^{(k)}\right)\delta\mathbf{a}^{(k)}-\mathbf{R}\right\|_{l_2}^2+\frac{\beta}{2}\left\|\mathbf{a}^{(k)}+\delta\mathbf{a}^{(k)}\right\|_{l_2}^2. \tag{41}$$

If the second term in (41) is replaced by $\frac{\beta}{2}\left\|\delta\mathbf{a}^{(k)}\right\|_{l_2}^2$ i.e. we set $\mathbf{a}^{(k)} = \mathbf{0}$, we recover the old LM perturbation (18):

$$\left[\mathbf{J}^\top\left(\mathbf{a}^{(k)}\right)\mathbf{J}\left(\mathbf{a}^{(k)}\right)+\beta\mathbf{I}\right]\delta\mathbf{a}^{(k)} = -g\left(\mathbf{a}^{(k)}\right), \tag{42}$$

which is a solution of the minimisation problem:

$$\min_{\delta\mathbf{a}\in\mathbb{R}^6}\frac{1}{2}\left\|F(\mathbf{a}^{(k)})+\mathbf{J}\left(\mathbf{a}^{(k)}\right)\delta\mathbf{a}^{(k)}-\mathbf{R}\right\|_{l_2}^2+\frac{\beta}{2}\left\|\delta\mathbf{a}^{(k)}\right\|_{l_2}^2. \tag{43}$$

Although the second term in (43) can be viewed as a regularizer, a Tikhonov-like term, for the perturbation $\delta\mathbf{a}$, the main problem with using this latter type of regularizers is that we cannot directly control the characteristics of the solution. In

other words, this approach does not take account into *a prior* information about the characteristics of solutions, which is the main task of regularization. In contrast, our RAR approach regularizes the current step $\left(\mathbf{a}^{(k)} + \delta\mathbf{a}\right)$ and so does control the characteristics of the solution.

6. A cooling method for the RAR parameter

As will be shown in Section 7, our RAR model is more robust than the standard affine model due to less demanding on good initial guesses. However the standard model does not need a regularization parameter β. Here we shall first present an algorithm to select the optimal β on the finest level L and then adapt the idea to a multiresolution setting to minimize the extra work.

There are many ways to select β — one option is to use the 'cooling' process (i.e. continuation) in an adaptive manner (see Haber and Oldenberg [8], Newman and Hoversten [24], Chen *et al.* [2], and Lelièvre and Oldenberg [17]). The basic idea is to start with a high initial value of β and then slowly reduce β in such a way that the solution obtained using it is an excellent starting point for the next, in order to decrease $\mathcal{J}_\beta$.

The initial β_1 is first estimated so that $\beta_1 \mathbf{H}_{\mathcal{S}}\left(\mathbf{a}_1^{(0)}\right)$ dominates the $\mathbf{J}^\top\left(\mathbf{a}_1^{(0)}\right)\mathbf{J}\left(\mathbf{a}_1^{(0)}\right)$ component in (38), where $\mathbf{a}_1^{(0)}$ is the initial guess solution. At the $(l+1)$th step we set

$$\beta_{l+1} = \gamma\beta_l \in [\beta_0, \beta_1], \tag{44}$$

where γ is a constant, usually chosen to be about 0.5, and β_0 is a small positive number, e.g. 5×10^{-5}. Subsequently, we apply β_{l+1} and the initial guess solution obtained by the previous iteration $\mathbf{a}_{l+1}^{(0)} = \mathbf{a}_l$ with the associated inner loop to obtain the minimum $\mathbf{a}_{l+1}$ within some tolerance. As mentioned in [8], since the functional $\mathcal{J}_\beta$ changes at each outer loop iteration, the demand of decreasing the value of the same functional is not reasonable so we impose the so-called *consistent condition* to ensure that the solution $\mathbf{a}_{l+1}$ and the parameter β_{l+1} are acceptable:

$$\mathcal{J}_{\beta_{l+1}}[\mathbf{a}_{l+1}] = \mathcal{D}[\mathbf{a}_{l+1}] + \beta_{l+1}\mathcal{S}[\mathbf{a}_{l+1}] < \mathcal{J}_{\beta_{l+1}}[\mathbf{a}_l] = \mathcal{D}[\mathbf{a}_l] + \beta_{l+1}\mathcal{S}[\mathbf{a}_l]. \tag{45}$$

If this condition is not satisfied, we increase γ (usually to 0.9) and re-start the step. Our experience suggests that the criterion given by

$$\frac{\|\mathbf{a}_{l+1} - \mathbf{a}_l\|}{\max\{\|\mathbf{a}_{l+1}\|, \|\mathbf{a}_l\|\}} < \delta \tag{46}$$

is suitable, where $\delta > 0$ is small (normally set to 5×10^{-4}). The process of solving the problem (36) for $\mathbf{a}$ with a given β (by the new RAR solver) will be denoted by

$$\mathbf{a}_\ell \leftarrow Solver\_RAR(\mathbf{T}_\ell, \mathbf{R}_\ell, \mathbf{a}_\ell, \beta, tol, IMAX)$$

for tolerance *tol* within the maximum number of iterations IMAX.

Finally, we summarize the unilevel cooling process as follows:

Algorithm 2 (Registration through cooling)**.**

$[\mathbf{a}^*, \beta^*] \leftarrow cooling(\mathbf{T}, \mathbf{R}, \mathbf{a}^{(0)}, \beta^{(0)})$

- Set $l = 1$, $\gamma = 0.5$, $\mathbf{a}_l = \mathbf{a}^{(0)}$, $\beta_0 = 5 \times 10^{-5}$ and $\beta_1 = \beta^{(0)}$. Set IMAX= 25 and $tol = 10^{-3}$.
- Outer iteration: For $l = 1, 2, 3, ...$
  - −1. Set $\beta_{l+1} = \gamma\beta_l$ in $[\beta_0, \beta_l]$
  - −2. Inner iteration: $\mathbf{a}_{new} \leftarrow Solver\_RAR(\mathbf{T}, \mathbf{R}, \mathbf{a}_l, \beta_{l+1}, tol)$.
  - −3. If $\mathcal{J}_{\beta_{l+1}}[\mathbf{a}_{new}] < \mathcal{J}_{\beta_{l+1}}[\mathbf{a}_l]$

−3.1. Set $\mathbf{a}_{l+1} = \mathbf{a}_{new}$, $\gamma = 0.5$, $l = l + 1$, and go to 4
Else
−3.2. Set $\gamma = 0.9$, and go to 4
−4. Check for convergence using the criterion (46).
If not satisfied, then return to 1, else, exit to the next step to stop.

- Set $\mathbf{a}^* = \mathbf{a}_{new}$ and $\beta^* = \beta_l$.

In the above algorithm, one notes that each minimization may not be solved exactly within IMAX iterations. Even so, the algorithm can be expensive for large images due to accumulated cost. Then our first robust algorithm will be the following.

Algorithm 3 (The basic RAR method)**.**

(1) *Input tol, given images* $\mathbf{T}$, $\mathbf{R}$. *Set* $\beta = 1$ *(optional). Set toln, MAXN.*
(2) *Obtain the optimal regularization parameter* β *(through cooling) via Algorithm 2:*

$$\left[\mathbf{a}^{(0)}, \beta\right] \leftarrow cooling(\mathbf{T}, \mathbf{R}, \mathbf{a}, \beta).$$

(3) *Solve the RAR problem (36) on the finest level using the found* β*:*

$$\mathbf{a} \leftarrow Solver\_RAR(\mathbf{T}, \mathbf{R}, \mathbf{a}^{(0)}, \beta, toln, MAXN).$$

In order to save computational work, we propose to use a hierarchy of L grids (with level L the finest and level 1 the coarsest one) as in Section 4.4. Firstly we shall seek the optimal β on the coarsest level 1 only and secondly we use the idea of Section 4.4 to provide finer level initial guesses. The whole procedure is summarized in Algorithm 4.

Algorithm 4 (Multilevel continuation for optimal β and reliable initial solution)**.**

$$[\mathbf{a}_\ell, \beta_\ell] \leftarrow RAR\_multiresolution(\mathbf{T}_\ell, \mathbf{R}_\ell, \mathbf{a}_\ell, \beta_\ell, \ell, tol)$$

- If $\ell = 1$
  - Set $\mathbf{a}_\ell = (1, 0, 0, 0, 1, 0)^\top$ or use Methods 1-3 in Section 4 to work out an initial $\mathbf{a}_\ell$
  - $\beta_\ell = C$ [Here $C > 0$ should be large enough e.g. $C = 1000$]
  - $[\mathbf{a}_\ell, \beta_\ell] \leftarrow cooling(\mathbf{T}_\ell, \mathbf{R}_\ell, \mathbf{a}_\ell, \beta_\ell)$
- Else
  - $\mathbf{T}_{\ell-1} = coarsen\,(\mathbf{T}_\ell)\,, \quad \mathbf{R}_{\ell-1} = coarsen\,(\mathbf{R}_\ell)$
  - $[\mathbf{a}_{\ell-1}, \beta_{\ell-1}] \leftarrow RAR\_multiresolution(\mathbf{T}_{\ell-1}, \mathbf{R}_{\ell-1}, \mathbf{a}_{\ell-1}, \beta_{\ell-1}, \ell-1, tol)$
  - $\mathbf{a}_\ell \leftarrow interpolate(\mathbf{a}_{\ell-1})$ as follows:
    $a_{i_\ell} = a_{i_{\ell-1}}$ for $i = 1, 2, 4, 5$ (the elements of the affine transformation matrix) and $a_{i_\ell} = 2a_{i_{\ell-1}}$ for $i = 3, 6$ (the elements of the translation matrix).
  - $\beta_\ell = 4\beta_{\ell-1}$ [Recall that $\beta_\ell = \bar{\beta} n_\ell^2$ and $n_\ell = 2n_{\ell-1}$].
  - $\mathbf{a}_\ell \leftarrow Solver\_RAR(\mathbf{T}_\ell, \mathbf{R}_\ell, \mathbf{a}_\ell, \beta_\ell, tol, \text{IMAX})$

Algorithm 5 (The refined RAR method)**.**

(1) *Input tol and set* $\mathbf{T}_L = \mathbf{T}$, $\mathbf{R}_L = \mathbf{R}$ *on the finest level. Set toln, MAXN.*
(2) *Obtain the optimal regularization parameter* β *(on the coarsest level* 1 *through cooling) and a good initial solution (through multi-resolution) via Algorithm 4:*

$$\left[\mathbf{a}^{(0)}, \beta\right] \leftarrow RAR\_multiresolution(\mathbf{T}_L, \mathbf{R}_L, \mathbf{a}_L, \beta_L, L).$$

(3) *Solve the RAR problem (36) on the finest level $\ell = L$ using the found β:*

$$\mathbf{a}_\ell \leftarrow Solver\_RAR(\mathbf{T}_\ell, \mathbf{R}_\ell, \mathbf{a}^{(0)}, \beta, toln, MAXN).$$

7. Numerical experiments

In this section, we shall give some results to illustrate the algorithms presented. Our first example (Example 5) is used to defend the integrity of our RAR method i.e. problems that possess genuinely large components in $\mathbf{a}$ are not penalized by our method (the regularization). Our second example (Example 6) will show that, for a nontrivial affine problem, the standard affine model even when Method 4 (Section 4.4) can fail to register properly while our RAR models (especially Algorithm 5) can register successfully. Our final example (Example 7) shows that, for the deformable problem (Example 4), our RAR method can provide a better initial solution than Method 4 (Section 4.4) which leads to even fewer number of FAS cycles by a deformable method (Section 3).

Example 5. *We consider a pair of synthetic images as in Figure 6 (a)−(b) with the images of size* 512×512. *Clearly one expects* $\mathbf{a}$ *will require large values.*

Using Algorithm 5 with $\mathcal{S}_1$, *we find that* $\mathbf{a}_{\mathcal{S}_1}$=$(0.2561, 0.4800, -134.4109, -0.2399, 0.8000, 275.9836)^\top$ *which is evidently not penalized by regularization. Similar solutions are obtained by* $\mathcal{S}_2, \mathcal{S}_3, \mathcal{S}_4$. *The successfully registered images using these* 4 *regularizers are respectively shown in Figure 6 (c), (d), (e) and (f). Here* $\mathbf{a}_{\mathcal{S}_1}$ *is the solution obtained from the regularizer* $\mathcal{S}_1$.

Example 6. *We consider an affine registration problem for a pair of MR images of a human head as in Figure 7 (a)−(b), where* $n_1 = n_2 = 256$. *We compare the GN and LM methods with Method* 4 *(Algorithm 1) with our RAR method (Algorithm 5).*

Since $\max\left\{ \|\mathbf{a}^* - \mathbf{a}_{\mathcal{S}_i}\|_{l_2} / \|\mathbf{a}^*\|_{l_2} \,\middle|\, i = 1, 2, 3, 4 \right\} = 0.0069$, *this means that our method converges to the true solution. Moreover, the registered images obtained from* 4 *different regularizers shown in 8* $(a) - (d)$ *are almost identical. Comparing those results obtained from the GN and LM methods (see Figure 7* $(c) - (d)$*) and our RAR method (see Figure 8* $(a) - (d)$*), one notes that the proposed latter method is more robust than the former methods.*

Example 7. *Finally, we re-solve Example 4 to show that Algorithm 5 is better than Algorithm 1 in affine pre-registration for the purpose of using a deformable model (via FAS algorithm).*

Here we show in Figure 9 (a)−(d) the four respective pre-registration images from our 4 *regularizers, and they appear identical. Indeed, using any of them to start FAS (Section 3) gives the same result as shown in Figure 9 (e)−(f) using (a)−(b) respectively. Moreover the details from Figure 9 (a)−(d) are visually more pleasing than Figure 5 (e)−(f) (especially at the upper region).*

To show the quantitative gain from using our Algorithm 5, we now present the comparable results in Table 1 for clarity, where 'Out.Iters' (same as l in step 3.1 of Algorithm 2) is the number of the outer iterations by Algorithm 3 and 'Avg.Iters' means the average of the number of inner iterations, given by

$$\text{Avg.Iters} = \frac{\text{The number of accumulated iterations by } \textit{Solver\_RAR} \text{ on the finest level}}{\text{The number of updates for parameter } \beta \text{ (via steps 3.1 and 3.2 in Algorithm 2)}}.$$

Clearly apart form the quality improvement over standard models (as illustrated before), much speed gain can be observed in Table 1 with our recommended Algorithm 5.

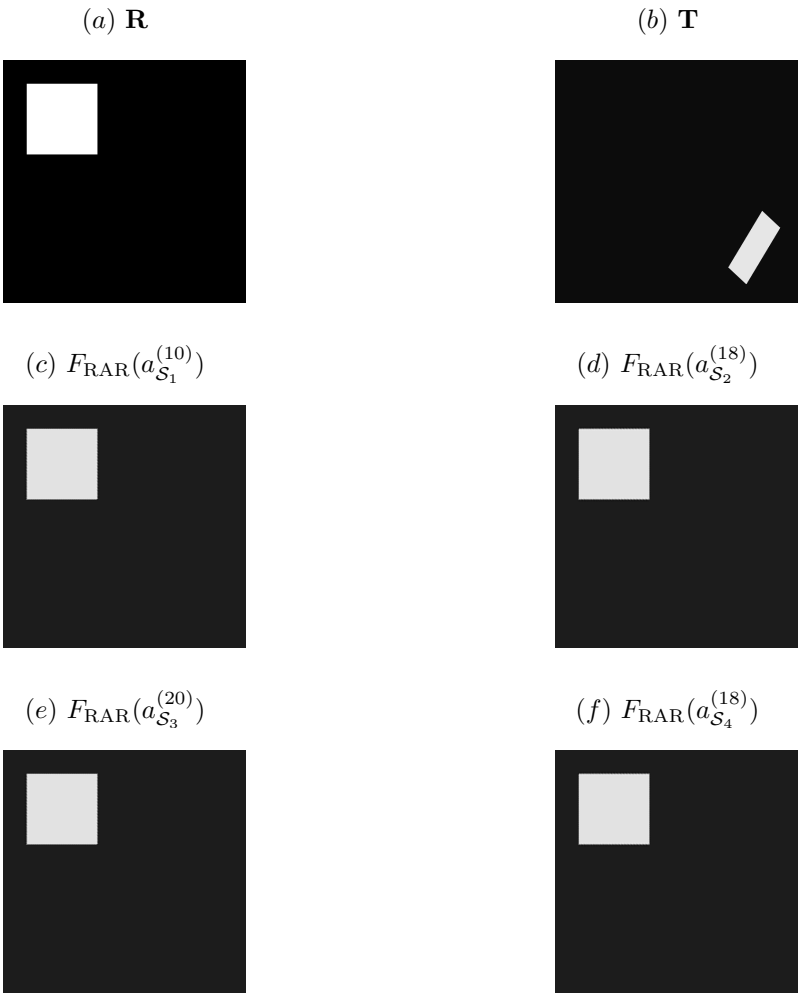


FIGURE 6. Example 5: Correct registration results (requiring large affine parameters) of a pair of synthetic images by our RAR model. The first row shows the reference (*a*) **R** and (*b*) the template **T**. The second and third rows show the registered images (c)−(f) from our 4 regularizers $\mathcal{S}_1 - \mathcal{S}_4$, respectively.

To summarize, in these and other tests, we have compared the performance of Algorithm 3 with 5. While both give comparable results, Algorithm 5 is much cheaper due to using a coarse level to work out β.

8. Conclusions

Parametric registration via a nonlinear least-square model offers a fast registration method. However the commonly used iterative methods of the GN (Gauss-Newton) and the LM (Levenberg-Marquardt) often have convergence difficulties, due to lack of good initial solutions, so the resulting nonlinear model is often not robust.

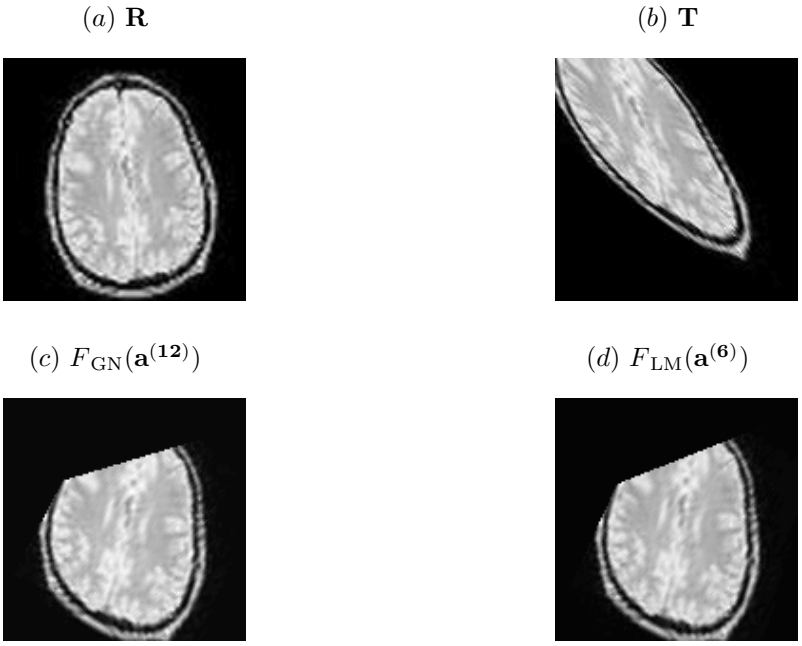


FIGURE 7. Example 6: Failed registration results of GN and LM with Method 4 (Algorithm 1). The first row shows the reference image: (*a*) **R** and the template image: (*b*) **T**. The second row presents the registered images: (*c*) $F_{\mathrm{GN}}(\mathbf{a}^{(12)})$ and (*d*) $F_{\mathrm{LM}}(\mathbf{a}^{(6)})$.

Example Number	Image size N	Algorithm 1 (GN) Iters/Ini.Cpu/Cpu	Algorithm 3 (S_1) Out.Iters/Avg.Iters/Cpu	Algorithm 5 (S_1) Iters/Ini.Cpu/Cpu
5	128^2	8/0.5/1.9	3/9/7.1	9/0.9/2.1
	256^2	7/1.0/6.9	3/9/22.8	7/1.5/7.0
	512^2	9/3.7/38.2	3/10/121.4	10/4.5/37.3
	1024^2	10/14.6/179.8	3/10/501.5	10/14.6/165.0
6	128^2	7/0.5/1.7	9/12/21.0	11/1.2/2.8
	256^2	12/1.1/14.3	9/12/207.6	11/1.4/10.7
	512^2	8/4.4/42.3	9/11/1006.0	10/8.3/46.9
	1024^2	24/19.3/558.2	9/11/3240.4	12/33.1/232.0

TABLE 1. Comparison of Algorithms $1, 3, 5$ using Examples $5 - 6$ with varying N.

In this paper, we first address the robustness issue of GN and LM methods for affine models by reviewing four methods for getting good initial guesses. It turns out that there are always difficult cases for which these initial guesses are not sufficient. Such cases include getting pre-registration images for deformable registration problems; we reviewed the diffusion based model and developed a FAS multigrid algorithm for testing purpose. Then, we propose a regularized affine registration (RAR) model that appears to be reliable and robust in solving i) the affine image registration problems ii) providing a good initial guess for deformable

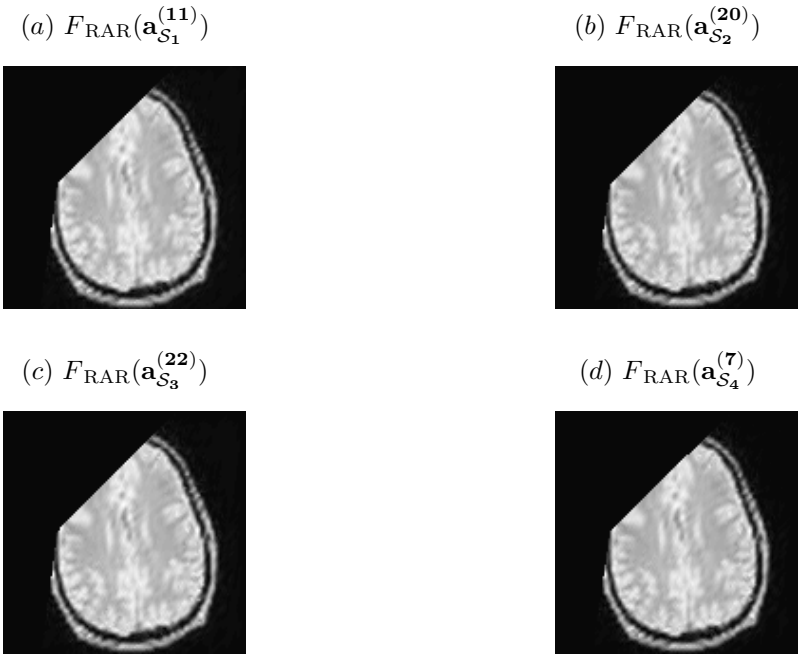


FIGURE 8. Example 6 re-solved: Correct registration results using our RAR method (Algorithm 5) with 4 regularizers $\mathcal{S}_1 - \mathcal{S}_4$, respectively shown in (a)-(d).

models. To find the optimal regularization parameter in an efficient way, we use a coarse level and a cooling idea, and combine with multi-resolution to initialize our RAR model. The resulting Algorithm 5 has been demonstrated in numerical experiments to be generally robust, for both affine problems and for pre-registration for deformable problems, without much increase in work.

Recently there was new work attempting to combine parametric and non-parametric models and we believe our idea of regularizing the parametric coefficients should be applicable there as well.

References

[1] T.F. Chan, and J.H. Shen. *Image Processing and Analysis - Variational, PDE, Wavelet, and Stochastic Methods*, SIAM Publications, Philadelphia, USA, 2005.

[2] J. Chen, E. Haber, and D.W. Oldenburg, "*3D numerical modelling and inversion of magnetometric resistivity data*", Geophys. J. Inter., Vol.**149** No.**3** (2002), pp.679–697.

[3] J.E. Dennis, and R.B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hal, Englewood Cliffs, 1983.

[4] A. Doicu, F. Schreier, and M. Hess, "*Iteratively regularized Gauss-Newton method for bound-constrained problem in atmospheric remote sensing*", Comp. Phys. Commu., Vol.**153** (2003), pp.59-65.

[5] A. Doicu, F. Schreier, and M. Hess, "*Iteratively regularized methods for remote sensing*", J. Quan. Spec., Vol.**83** (2004), pp.47-61.

[6] R. Fletcher, *Practical Methods of Optimization*, John Wiley and Sons, 1987.

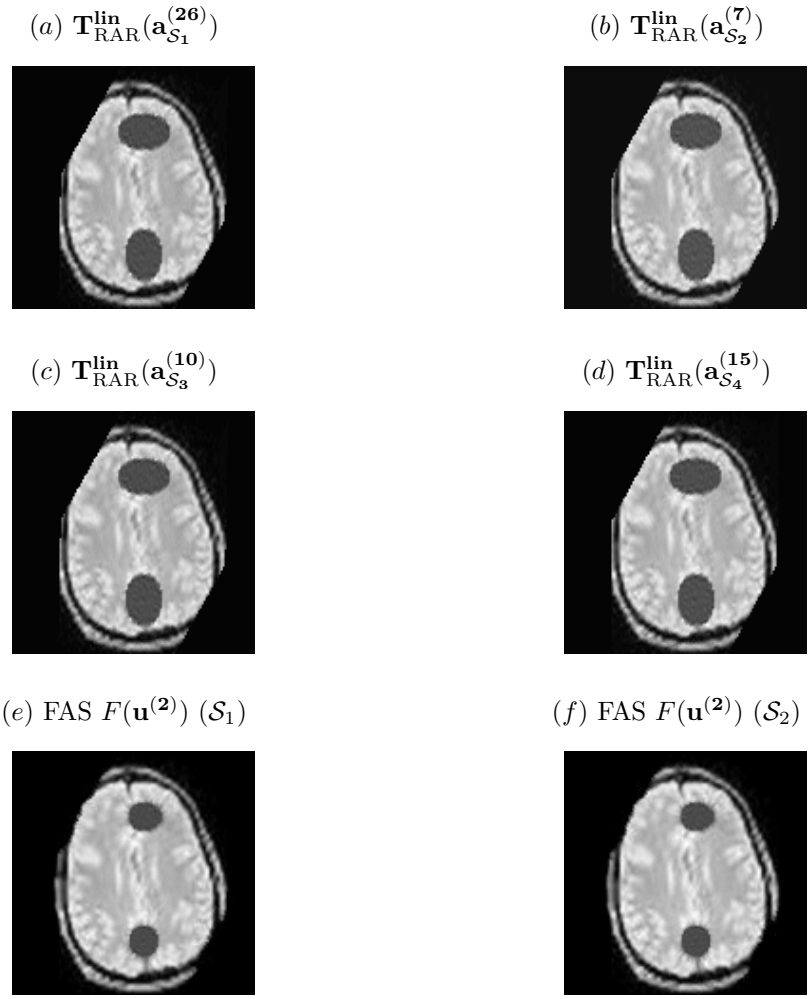


FIGURE 9. Example 4 re-solved and improved: The affine pre-registration steps (a)−(d) using Algorithm 5 with different regularizers $\mathcal{S}_1-\mathcal{S}_4$, respectively. The last row (e)-(f) shows the respective registered images by FAS method with (a)-(b) as initial solutions (using the (c)-(d) gives almost identical solutions). Clearly less FAS cycles (i.e. 2) are needed than before (i.e. 6).

[7] B. Fischer, and J. Modersitzki, "*Fast diffusion registration*", in M.Z. Nashed, O. Scherzer (eds), Inverse Problems, Image Analysis, and Medical Imaging (New Orleans, LA, 2001), , Amer. Math. Soc., Providence, RI. Contemp. Math., Vol.**313** (2002), pp.117-129.

[8] E. Haber, and D.W. Oldenburg, "*A GCV based method for nonlinear ill-posed problems*", Computational Geosciences, Vol.**4** No.**1** (2000), pp.41-63.

[9] J.V. Hajnal, D.L.G. Hill, and D. Hawkes, Medical Image Registration, The Biomedical Engineering Series, CRC Press, 2001.

[10] M. Hanke, "*A regularizing Levenberg-Marquart scheme, with applications to inverse groundwater filtration problems*", Inv. Problems, Vol.**13** (1997), pp.79-95.
[11] S. Henn, "*A Levenberg-Marquart scheme for nonlinear image registration*", BIT Numer. Math., Vol.**43** (2003), pp.743-759.
[12] S. Henn, and K. Witsch, "*Iterative multigrid regularization techniques for image matching*", SIAM J. Sci. Comput., Vol. **23** No.**4** (2001), pp.1077-1093.
[13] D.L.G. Hill, P.G. Batchelor, M. Holden, and D.J. Hawkes, "*Medical image registration*", Phys. Med. Biol., Vol. **46** (2001), pp.1-45.
[14] M. Jenkinson, P. Bannister, M. Brady, and S. Smith, "*Improved optimisation for the robust and accurate linear registration and motion correction of brain images*", NeuroImage, Vol. **17** No.**2** (2002), pp.825-841.
[15] M. Jenkinson, and S. Smith, "*A global optimisation method for robust affine registration of brain images*", Med. Image Anal., Vol. **5** (2001), pp.143-156.
[16] C.T. Kelly, *Iterative Methods for Optimization* (Frontiers in Applied Mathematics), SIAM Publications, Philadelphia, USA, 1999.
[17] P.G. Lelièvre, and D.W. Oldenburg, "*Magnetic forward modelling and inversion for high susceptibility*", Geophys. J. Inter., Vol.**166** No.**1** (2006), pp.76-90.
[18] W. Lu, M. Chen, G.H. Olivera, K.J. Ruchala, and T.R. Mackie, "*Fast free-form deformable registration via calculus of variation*", Phys. Med. Biol., Vol.**49** No.**14** (2004), pp.3067-3087.
[19] T. Lu, P. Neittaanmäki, and X.-C. Tai, "*A parallel splitting up method and its application to Navier-Stokes equations*", Appl. Math. Lett., Vol.**4** (1991), pp.25-29.
[20] L. Lucchese, S. Leorin, and G.M. Cortrlazzo, "*Estimation of 2D affine transformations through polar curve matching and its application to image mosaicking and remote-sensing data registration*", IEEE Trans. Image Proc., Vol.**15** No.**10** (2006), pp.3008-3018.
[21] F. Maes, D. Vandermeulen, and P. Suetens, "*Comparative evaluation of multiresolution optimization strtegies for multimodal image registration by maximization of mutual information*", Med. Image Anal., Vol.**3** No.**4** (1999), pp.373-386.
[22] J.B.A. Maintz, and M.A. Viergever, "*A survey of medical image registration*", Med. Image Anal., Vol. **2** No.**1**(1998), pp.1-36.
[23] J. Modersitzki, *Numerical Methods for Image Registration*, Oxford University Press, 2004.
[24] G. A Newman, and G M. Hoversten, "*Solution strategies for two- and three-dimensional electromagnetic inverse problems*", Inv. Problems, Vol.**16**(2000), pp.1357–1375.
[25] J. Nocedal, and S.J. Wright, *Numerical Optimization*, Springer-Verlag, NewYork, 1999.
[26] R.L. Parker, *Geophysical Inverse Theory*, Princeton University Press, 1994.
[27] J.F. Ralph, *Automatic Image Matching (AIM) GUI Toolset*, Department of Electrical Engineering and Electronics, The University of Liverpool, 2006.
[28] D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach, and D.J. Hawkes, "*Nonrigid registration using free-form deformations: application to breast MR images*", IEEE Trans. Med. Imaging, Vol.**18** No.**8** (1999), pp.712-721.
[29] O. Schmitt, J. Modersitzki, S. Heldmann, S. Wirtz, and B. Fischer,"*Image registration of sectioned brains*", Int. J. Comp. Vision., Vol **73** No.**1** (2007), pp.5-39.
[30] E. Spedicato, and M.T. Vespucci, "*Numerical experiments with variations of the Gauss-Newton algorithm for nonlinear least squares*", J. Opt. Theory Appl., Vol **57** No.**2** (1988), pp.323-339.
[31] C. Studholme, D.L.G. Hill, and D.J. Hawkes, "*Automated 3-D registration of MR and CT images of the head*", Med. Image. Anal., Vol. **1** No.**2** (1996), pp.163-175.
[32] P. Thévenaz, U. E. Ruttimann, and M. Unser, "*A pyramid approach to subpixel registration based on intensity*", IEEE Trans. Image Proc., Vol.**7** No.**1** (1998), pp.27-41.
[33] U. Trottenberg, C. Oosterlee, and A. Schuller. *Multigrid.* Academic Press, 2001.

[34] C.R. Vogel, *Computational Methods for Inverse Problems* (Frontiers in Applied Mathematics), SIAM Publications, Philadelphia, USA, 2002.
[35] J. Weickert, B. M. ter Haar Romeny, and M. A. Viergever, "*Efficient and reliable schemes for nonlinear diffusion filtering*", IEEE Trans. Image Proc., Vol.**7** (1998), pp.398-410.
[36] M. Xia, and B. Liu, "*Image registration by 'super-curves'*", IEEE Trans. Image Proc., Vol.**13** No.**5** (2004), pp.720-732.
[37] Y.X. Yuan, and W.Y. Sun, *Optimization Theories and Methods*, Science Press, Beijing, China, 1997.
[38] P. Zhilkin, and M.E. Alexander, "*Affine registration: a comparison of several programs*", J. Magn. Reson. Imaging, Vol.**22** (2004), pp.55-66.

Centre for Mathematical Imaging Techniques and Department of Mathematical Sciences, The University of Liverpool, Peach Street, Liverpool L69 7ZL, United Kingdom.
E-mail: cnoppado@lliverpool.ac.uk and k.chen@liverpool.ac.uk []
URL: http://www.liv.ac.uk/www/cmit/