Link to formal publication Medical Image Analysis (Elsevier): https://doi.org/10.1016/j.media.2016.07.003

Stochastic Optimization with Randomized Smoothing for Image Registration

Wei Sun[a,d,*], Dirk H.J. Poot[a,b], Ihor Smal[a], Xuan Yang[c], Wiro J. Niessen[a,b], Stefan Klein[a]

[a] *Biomedical Imaging Group Rotterdam, Departments of Radiology and Medical Informatics, Erasmus MC, Rotterdam, The Netherlands*
[b] *Department of Image Science and Technology, Faculty of Applied Sciences, Delft University of Technology, Delft, The Netherlands*
[c] *College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, Guangdong, China*
[d] *Laboratory of Neuro Imaging (LONI), USC Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California, Los Angeles, USA*

Abstract

Image registration is typically formulated as an optimization process, which aims to find the optimal transformation parameters of a given transformation model by minimizing a cost function. Local minima may exist in the optimization landscape, which could hamper the optimization process. To eliminate local minima, smoothing the cost function would be desirable. In this paper, we investigate the use of a randomized smoothing (RS) technique for stochastic gradient descent (SGD) optimization, to effectively smooth the cost function. In this approach, Gaussian noise is added to the transformation parameters prior to computing the cost function gradient in each iteration of the SGD optimizer. The approach is suitable for both rigid and nonrigid registrations. Experiments on synthetic images, cell images, public CT lung data, and public MR brain data demonstrate the effectiveness of the novel RS technique in terms of registration accuracy and robustness.

Keywords:
Image registration, Local minima, Stochastic gradient descent, Randomized smoothing

1. Introduction

Image registration is often described as an ill-posed optimization problem (Fischer and Modersitzki, 2008). To solve a registration task, a cost function is defined, and then it is minimized by a numerical optimization routine. During the optimization process, local minima in the cost function may trap the optimization to a sub-optimal solution. This is a common cause of misregistration.

The optimization landscape is fully defined by the constituents of the cost function, such as input image data, similarity metric, transformation model and interpolation method. Jenkinson et al. (2002) classified the local minima that commonly occur for the cost functions into two types: large-scale basins and small-scale dips. They concluded that the large-scale basins are responsible for large misregistrations because they are often relatively far from the global minimum. For example, neighboring structures with similar appearance could cause a wrong match, especially in case of a large initial misalignment. The small-scale dips occur more often and could cause the registration to get stuck at any stage. A well-known example of such small scale dips are the local minima induced by interpolation artifacts (Pluim et al., 2000; Likar and Pernuš, 2001; Tsao, 2003; Aganj et al., 2013).

In intensity-based image registration a multiresolution Gaussian stack of image data is often used to reduce local minima (Lester and Arridge, 1999; Sun et al., 2013). Typically, the capture range of the optimal solution is enlarged by using a multiresolution strategy. However, Jenkinson and Smith (2001) argued that the traditional multiresolution strategy cannot solve all possible local minima in image registration. Even though multiresolution approaches reduce the effect of local minima, on a certain scale level the convexity of the cost function is not guaranteed in practice. In addition, the capture ranges of adjacent resolution levels should be close enough without any barrier in between them. In their research (Jenkinson and Smith, 2001), the authors proposed a multi-start algorithm, using m perturbations of the initial (affine) transformation parameters, to achieve a global optimization. They found

*Corresponding author. Address: Biomedical Imaging Group Rotterdam, Erasmus Medical Center, 3015 GE Rotterdam, the Netherlands. Tel.: +31 107044078.
Email address: sunwei@ieee.org (Wei Sun)

that these perturbations correct for the majority of misregistrations. However, the computational cost for this multi-start algorithm is increased by the factor m. In addition, this method becomes computationally impractical for high dimensional, nonrigid transformation models, since very high m would be needed. In their later research (Jenkinson et al., 2002), two strategies are proposed to reduce the effects of both large- and small-scale local minima separately. For large-scale local minima, they proposed to use a hybrid global-local optimization method that exploits prior knowledge about the linear transformation parameters. For small-scale local minima, the cost function is apodized by reweighting the contribution of intensity information. However, their method can only be utilized for linear transformation models rather than more complicated nonlinear cases. In a number of papers (Likar and Pernuš, 2001; Tsao, 2003; Aganj et al., 2013; Thévenaz et al., 2006; Klein et al., 2010) random coordinate (re)sampling was proposed as a method for removing small-scale local minima due to interpolation artifacts. These techniques can be applied both to linear and nonlinear registration, but they do not address the large-scale local minima and non-interpolation-related small-scale local minima. Another often used approach in nonrigid registration methods is to add a regularization term to the cost function, in order to promote smooth deformations, see for example (Rueckert et al., 1999; Avants et al., 2008; Glocker et al., 2008). However, smoothness of the deformation field does not imply that local minima in the cost function are eliminated. Even with linear transformation models (i.e., rigid and affine), which are smooth by definition, local minima pose a challenge. So, although regularization terms may in practice suppress the effect of some local minima, they do not directly target that issue. Moreover, these methods are only devised for nonrigid transformation models. In this paper we address the issue of local minima directly and our approach is applicable to any transformation model.

In our research, we try to tackle the problem by *smoothing the optimization landscape*. Local minima in the optimization landscape could potentially be eliminated by smoothing the cost function. Multiresolution strategies indirectly achieve some sort of cost function smoothing by blurring the images, but in our approach we investigate to directly smooth the cost function. Conceptually, a straightforward way to smooth the cost function would be by convolution with a smoothing kernel in the transformation parameter space. However, a major challenge here is the high dimensionality of the parameter space. To address this challenge, we will borrow an idea from the literature on numerical optimization.

In the field of optimization, generic techniques for smoothing the cost function have been investigated (Nesterov, 2005; Beck and Teboulle, 2012; Duchi et al., 2012). Research on this topic has mainly focused on non-smooth, not everywhere differentiable cost functions. Nesterov (2005) constructed a function with Lipschitz-continuous gradient to approximate the original non-smooth cost function. Instead of approximating the entire non-smooth cost function, Beck and Teboulle (2012) proposed to only smooth the non-smooth component of a cost function. Duchi et al. (2012) proposed a randomized smoothing (RS) algorithm for stochastic optimization problems where only a stochastic (noisy) measurement of the cost function gradient is available. In the RS algorithm, random variables are drawn from a Gaussian or uniform distribution, and these variables are added to the estimated parameters during optimization. The underlying idea is a convolution-based smoothing technique in which a Gaussian or uniform distribution is convolved with the original cost function.

Inspired by Duchi et al. (2012), we propose to inject Gaussian random noise into the transformation parameters during the optimization process of image registration. In contrast to the costly multi-start algorithm by Jenkinson and Smith (2001), which applies multiple perturbations to the initial transformation parameters, we apply one (or just a few) random perturbation(s) to the transformation parameters in each iteration. This RS technique is combined with a stochastic gradient descent optimizer (Klein et al., 2009b), which can naturally deal with the random perturbations. The RS method could be seen as a computationally efficient way that implicitly smoothes the cost function, thus eliminating local minima in the optimization landscape. We evaluate this assertion in experiments on synthetic data, where we compare the proposed RS approach with a method that explicitly smoothes the cost function in a brute-force way. Since it is a generic technique, the proposed RS registration method can in principle be used in combination with any transformation model. In this study, translation, rigid, affine and nonrigid B-spline transformation models are used. Experiments on two-dimensional artificial images, two-dimensional cell images, three-dimensional lung CT, and brain MRI scans were carried out to evaluate the effectiveness of the RS registration approach.

2. Method

2.1. Stochastic optimization for image registration

Let $F(\boldsymbol{x}) : \Omega_F \subset \mathbb{R}^D \rightarrow \mathbb{R}$ and $M(\boldsymbol{x}) : \Omega_M \subset \mathbb{R}^D \rightarrow \mathbb{R}$ denote the D-dimensional fixed and moving images where $\boldsymbol{x}$ represents an image coordinate, and Ω_F and Ω_M are the fixed and moving image domains, respectively. Assume $\mathbf{T}(\boldsymbol{\mu}, \boldsymbol{x}) : \mathbb{R}^P \times \Omega_F \rightarrow \Omega_M$ represents a coordinate transformation where $\boldsymbol{\mu} \in \mathbb{R}^P$ is the parameter vector of the transformation model. $\mathbf{T}(\boldsymbol{\mu}, \boldsymbol{x})$ could be a translation, rigid, affine or nonrigid (e.g., B-spline) transformation model. Then, the registration problem is formulated as:

$$\hat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} C\left(\boldsymbol{\mu}, \Omega_F\right), \tag{1}$$

where $C(\boldsymbol{\mu}, \Omega_F)$ measures the dissimilarity between the original fixed image $F(\boldsymbol{x})$ and the deformed moving image $M(\mathbf{T}(\boldsymbol{\mu}, \boldsymbol{x}))$ on the domain $\boldsymbol{x} \in \Omega_F$. Examples of C are the sum of squared differences (SSD), normalized correlation coefficient (NCC), and mutual information (Viola and Wells III, 1997; Maes et al., 1997). For example, the cost function C of SSD is defined as:

$$C\left(\boldsymbol{\mu}, \Omega_F\right) = \frac{1}{|\Omega_F|} \sum_{\boldsymbol{x}_i \in \Omega_F} \left(F(\boldsymbol{x}_i) - M\left(\mathbf{T}(\boldsymbol{\mu}, \boldsymbol{x}_i)\right)\right)^2. \tag{2}$$

Since Eq. (1) has no closed-form solution in many realistic cases, an iterative optimization strategy is utilized to determine the optimal set of parameters $\hat{\boldsymbol{\mu}}$. Well-known instances of such optimizers are gradient descent, quasi-Newton, and nonlinear conjugate gradient (Maes et al., 1999). Viola and Wells III (1997) proposed to use SGD optimization for more efficient rigid image registration. A comparison of different optimizers is reported in Klein et al. (2007) indicating that the SGD optimizer is a competitive alternative to deterministic algorithms in nonrigid registration problems as well. SGD optimization is based on the following iterative update strategy:

$$\boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k - a_k \tilde{\boldsymbol{g}}_k, \tag{3}$$

where $\tilde{\boldsymbol{g}}_k$ is a stochastic approximation of the cost function derivative $\partial C / \partial \boldsymbol{\mu}$, evaluated at the current recovered transformation parameters $\boldsymbol{\mu}_k$, and a_k is a scalar gain factor that controls the step size along $\tilde{\boldsymbol{g}}_k$. Convergence in SGD methods can be achieved by decaying a_k according to a pre-defined function. In Klein et al. (2009b) an adaptive strategy for setting a_k was proposed. This adaptive stochastic gradient descent (ASGD) method is used in this work. In previous studies, (A)SGD optimization has been successfully applied in many image registration tasks, such as Bhagalia et al. (2009); Metz et al. (2011); Murphy et al. (2011); Smal et al. (2012); Sun et al. (2013).

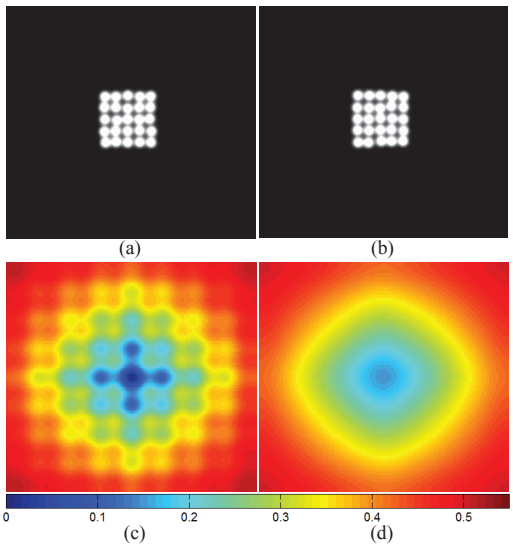

Figure 1: Smoothing an optimization landscape with local minima: (a) fixed image; (b) moving image; (c) optimization landscape of the translation transformation with X-translation $t_x \in [-80, 80]$mm and Y-translation $t_y \in [-80, 80]$mm; (d) smoothed optimization landscape by convolving with a Gaussian kernel using $\sigma = 10$mm along t_x and t_y.

In Klein et al. (2007), the stochastic approximation $\tilde{\boldsymbol{g}}_k$ was calculated by evaluating $\partial C / \partial \boldsymbol{\mu}$ on a small random subset $\widetilde{\Omega}_F^k \subset \Omega_F$ of image samples, thus reducing the computation time per iteration. This subset $\widetilde{\Omega}_F^k$ should be randomly refreshed in each iteration k, to make the approximation stochastic. We thus can write:

$$\tilde{\boldsymbol{g}}_k = \tilde{\boldsymbol{g}}(\boldsymbol{\mu}_k, \widetilde{\Omega}_F^k) = \frac{\partial C}{\partial \boldsymbol{\mu}}(\boldsymbol{\mu}_k, \widetilde{\Omega}_F^k) \approx \frac{\partial C}{\partial \boldsymbol{\mu}}(\boldsymbol{\mu}_k, \Omega_F). \tag{4}$$

For example, if we choose SSD as cost function C (see Eq. (2)), $\tilde{\boldsymbol{g}}_k$ is computed as:

$$\tilde{\boldsymbol{g}}_k = \frac{2}{|\widetilde{\Omega}_F^k|} \sum_{\boldsymbol{x}_i \in \widetilde{\Omega}_F^k} \left(F(\boldsymbol{x}_i) - M\left(\mathbf{T}(\boldsymbol{\mu}_k, \boldsymbol{x}_i)\right)\right) \tag{5}$$
$$\times \left(\left. \frac{\partial \mathbf{T}}{\partial \boldsymbol{\mu}} \right|_{(\boldsymbol{\mu}_k, \boldsymbol{x}_i)} \right)^T \left(\left. \frac{\partial M}{\partial \boldsymbol{x}} \right|_{\mathbf{T}(\boldsymbol{\mu}_k, \boldsymbol{x}_i)} \right).$$

2.2. Randomized smoothing

To avoid local minima and attain the global optimum, smoothing the cost function of image registration would be an attractive solution. To smooth the cost function a straightforward approach would be to convolve it with a

probability density function (PDF) $\zeta(\boldsymbol{\psi})$,

$$\bar{C}(\boldsymbol{\mu}, \Omega_F) = \int_{\boldsymbol{\psi} \in \mathbb{R}^P} C(\boldsymbol{\mu} - \boldsymbol{\psi}, \Omega_F)\, \zeta(\boldsymbol{\psi}) d\boldsymbol{\psi} \qquad (6)$$
$$= \mathbb{E}\left[C(\boldsymbol{\mu} + \boldsymbol{\Psi}, \Omega_F)\right],$$

where $\boldsymbol{\Psi}$ is a zero-mean random multivariate variable according to symmetric PDF $\zeta(\boldsymbol{\psi})$, and $\mathbb{E}[C(\boldsymbol{\mu} + \boldsymbol{\Psi}, \Omega_F)]$ represents the expectation of $C(\boldsymbol{\mu} + \boldsymbol{\Psi}, \Omega_F)$ given $\zeta(\boldsymbol{\psi})$.

Figure 1 shows an example of smoothing the cost function by using Eq. (6). Figures 1 (a) and (b) show two 2D artificial images with 25 Gaussian blobs randomly shifted from regular grid positions. The pixel size and dimension of each image are 1×1mm and 400×400 pixels. Figure 1 (c) plots the SSD (cost function) values between the fixed image (Figure 1 (a)) and the moving image (Figure 1 (b)) as function of X-translation $t_x \in [-80, 80]$mm and Y-translation $t_y \in [-80, 80]$mm. Due to the image content, the optimization landscape contains many local minima, and a typical gradient descent based optimizer could be trapped in local minima. Figure 1 (d) shows a smoothed optimization landscape obtained using Eq. (6) by convolving the original cost function with a Gaussian kernel using $\sigma = 10$mm along t_x and t_y. It can be observed that the local minima in Figure 1 (c) are effectively removed by using the expectation of cost function.

Unfortunately, it is hard to evaluate the integral in Eq. (6) in practice since the number of transformation parameters is usually large. To address this, we propose to use a randomized smoothing (RS) technique. The RS technique enables us to minimize $\bar{C}(\boldsymbol{\mu}, \Omega_F)$ without actually calculating the integral in Eq. (6). The optimization with the RS technique could be considered a special case of a SGD optimization algorithm. In general, SGD algorithms are exactly meant for the task of minimizing an expected value of a cost function in cases where this expected value cannot be efficiently evaluated in practice. First, observe that the expected value in Eq. (6) could be approximated by the following finite sum by using the RS technique:

$$\mathbb{E}\left[C(\boldsymbol{\mu} + \boldsymbol{\Psi}, \Omega_F)\right] \approx \frac{1}{Q} \sum_{q=1}^{Q} C(\boldsymbol{\mu} + \boldsymbol{\Psi}_q, \Omega_F), \qquad (7)$$

with $\boldsymbol{\Psi}_q \sim \zeta(\boldsymbol{\psi})$.

Therefore, the same approximation can be made for the gradient of the cost function. The RS algorithm is nothing more than a SGD algorithm, i.e., Eq. (3), where the stochastic gradient $\tilde{\boldsymbol{g}}_k$ is re-defined as:

$$\tilde{\boldsymbol{g}}_k = \frac{1}{Q} \sum_{q=1}^{Q} \tilde{\boldsymbol{g}}(\boldsymbol{\mu}_k + \boldsymbol{\Psi}_q, \widetilde{\Omega}_F^k). \qquad (8)$$

In each iteration k, the derivative of the original cost function is thus queried several (Q) times with different realizations of the random variable $\boldsymbol{\Psi}_q$. As in the original SGD algorithm, these derivatives are computed on a random subset $\widetilde{\Omega}_F^k$. Thanks to the properties of the SGD algorithm, the iterative scheme Eq. (3) with this choice for $\tilde{\boldsymbol{g}}_k$ will converge to the minimizer of Eq. (6). A theoretical analysis of the convergence of the RS technique is provided in Duchi et al. (2012). Increasing Q reduces the variance of $\tilde{\boldsymbol{g}}_k$ and may thus accelerate convergence, but increases the computational complexity per iteration. Choosing the value of Q is an application-dependent trade-off. Once the optimizer is sufficiently close to the desired global optimum, it might be desirable to reduce the amount of smoothing. It is thus proposed to let $\boldsymbol{\Psi}_q \sim \zeta(\boldsymbol{\psi}/h_k)/h_k$, where h_k is some nonincreasing function that controls the amplitude of noise. The choice of h_k controls the trade-off between the effective amount of cost function smoothing (i.e., the ability to jump out of local minima) and the variance of the estimate $\boldsymbol{\mu}_k$.

It is worth noting that the RS algorithm treats the cost function as a black box, which means that it does not require modification of the cost function itself (Nemirovsky and Yudin, 1983). It could be considered a "plug-in" technique, which only requires access to a function that evaluates the cost function gradient for a specified transformation parameter vector. This makes the technique potentially applicable to many registration methods.

Theoretically, we can use any smoothing PDF $\zeta(\boldsymbol{\psi})$, e.g., uniform or Gaussian. In this work, we use a Gaussian distribution with *mean* = 0 and user-defined σ. The Gaussian distribution was truncated at $\pm 2\sigma$. The traditional approach without added noise can be considered adding Gaussian noise with *mean* = 0 and $\sigma = 0$ (i.e., a Dirac delta distribution). Both approaches with ($\sigma > 0$) and without ($\sigma = 0$) the RS technique will be evaluated and compared in the experiments. The function h_k that controls the amplitude of noise is chosen as an exponentially decreasing function $h_k = e^{-k/(\lambda \cdot K)}$ where λ represents the decrease rate of h_k, K is the number of iterations and $k \leq K$.

Algorithm 1 presents the registration algorithm with the RS technique, assuming SSD as a similarity measure for the sake of clarity (similarity measures like MI involve an additional loop over the samples $\boldsymbol{x}_i$, to compute the joint histogram, but the principle is the same).

2.3. Transformation models

In the RS technique, random noise is directly added to transformation parameters. Therefore, RS can in prin-

Input: $F \leftarrow$ fixed image, $M \leftarrow$ moving image,
$K \leftarrow$ number of iterations, $S \leftarrow$ number of samples $|\widetilde{\Omega}_F|$,
$Q \leftarrow$ number of queries and
$\zeta(\boldsymbol{\psi}) \leftarrow$ smoothing PDF (e.g., Gaussian)
Output: Estimated transformation parameters $\hat{\boldsymbol{\mu}}$

1 Initialize transformation parameters $\boldsymbol{\mu}_k \leftarrow \mathbf{0}$
2 **for** $k \leftarrow 1$ *to* K **do**
3 Initialize random samples $\widetilde{\Omega}_F^k = \{\boldsymbol{x}_1 \dots \boldsymbol{x}_S\}$ according to uniform distribution, $\tilde{\boldsymbol{g}}_k = \mathbf{0}$, step size a_k, and noise level h_k
4 **for** $q \leftarrow 1$ *to* Q **do**
5 Draw random noise $\boldsymbol{\Psi}_q \sim \zeta(\boldsymbol{\psi}/h_k)/h_k$
6 $\tilde{\boldsymbol{\mu}} \leftarrow \boldsymbol{\mu}_k + \boldsymbol{\Psi}_q$
7 **for** $\boldsymbol{x} \leftarrow \boldsymbol{x}_1$ *to* $\boldsymbol{x}_S$ **do**
8 Evaluate $F(\boldsymbol{x})$, and $\mathbf{y} \leftarrow \mathbf{T}(\tilde{\boldsymbol{\mu}}, \boldsymbol{x})$
9 Interpolate moving image value $M(\mathbf{y})$ and gradient $\nabla M(\mathbf{y})$
10 Calculate $\left.\frac{\partial \mathbf{T}}{\partial \boldsymbol{\mu}}\right|_{(\tilde{\boldsymbol{\mu}}, \boldsymbol{x})}$
11 Calculate contribution to $\tilde{\boldsymbol{g}}(\boldsymbol{\mu}_k + \boldsymbol{\Psi}_q, \widetilde{\Omega}_F^k)$ using the results of steps 8-10
12 **end**
13 **end**
14 $\tilde{\boldsymbol{g}}_k = \frac{1}{Q} \sum_{q=1}^{Q} \tilde{\boldsymbol{g}}(\boldsymbol{\mu}_k + \boldsymbol{\Psi}_q, \widetilde{\Omega}_F^k)$
15 Update transformation parameters $\boldsymbol{\mu}_{k+1} \leftarrow \boldsymbol{\mu}_k - a_k \tilde{\boldsymbol{g}}_k$
16 **end**
17 $\hat{\boldsymbol{\mu}} \leftarrow \boldsymbol{\mu}_K$
18 **return** $\hat{\boldsymbol{\mu}}$

Algorithm 1: Proposed registration method.

ciple be used in combination with any transformation model. We evaluate the RS technique with global translation, rigid, and affine transformations, and with B-spline based free-form deformation (FFD) models.

2.3.1. Translation transformation

Translation transformation is defined as:

$$\mathbf{T}(\boldsymbol{\mu}, \boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{\mu}, \tag{9}$$

where $\boldsymbol{\mu}$ represents the translation vector with length D.

2.3.2. Rigid transformation

A rigid transformation is defined as:

$$\mathbf{T}(\boldsymbol{\mu}, \boldsymbol{x}) = \boldsymbol{R}(\boldsymbol{x} - \boldsymbol{c}) + \boldsymbol{t} + \boldsymbol{c}, \tag{10}$$

where $\boldsymbol{R}$ represents a rotation matrix, $\boldsymbol{c}$ is the center of rotation, and $\boldsymbol{t}$ is the vector of translations. The transformation of the rigid model is parameterized using $\boldsymbol{\mu} = (\boldsymbol{\theta}^T, \boldsymbol{t}^T)^T$ where $\boldsymbol{\theta}$ represents the vector of Euler angles. For example transformation parameters $\boldsymbol{\mu}$ are expressed as $(\theta, t_x, t_y)^T$ in 2D registration. Since the Euler angles can have an entirely different range than the translations, the following reparameterization in ASGD (Klein et al., 2009b) is used:

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\Gamma} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{t} \end{bmatrix}, \tag{11}$$

where $\boldsymbol{\Gamma}$ is a diagonal scaling matrix with the diagonal element:

$$\Gamma_{ii} = \left(\int_{\Omega_F} \left\| \frac{\partial \mathbf{T}}{\partial \theta_i}(\boldsymbol{\mu}_0, \boldsymbol{x}) \right\|^2 d\boldsymbol{x} \Big/ \int_{\Omega_F} d\boldsymbol{x} \right)^{-\frac{1}{2}}. \tag{12}$$

In this way, the rotation parameters $\boldsymbol{\theta}$ are scaled by the average voxel displacement caused by a small change of the Euler angle. Through the above normalizing process, it is meaningful to add Gaussian noise to different transformation parameters regardless their ranges.

2.3.3. Affine transformation

An affine transformation can be formulated as:

$$\mathbf{T}(\boldsymbol{\mu}, \boldsymbol{x}) = \boldsymbol{A}(\boldsymbol{x} - \boldsymbol{c}) + \boldsymbol{t} + \boldsymbol{c}, \tag{13}$$

where $\boldsymbol{A}$ is an affine matrix, $\boldsymbol{c}$ is also the center of rotation, and $\boldsymbol{t}$ represents the vector of translations. The parameter vector $\boldsymbol{\mu}$ is formed by affine matrix elements and translation vector. For example, in 2D registration this gives a vector of length 6: $\boldsymbol{\mu} = (a_{11}, a_{12}, a_{21}, a_{22}, t_x, t_y)^T$. The parameters are normalized using a similar procedure as described in Eqs. (11) and (12).

2.3.4. B-spline based FFD transformation

The traditional FFD transformation model (Rueckert et al., 1999) is defined as:

$$\mathbf{T}(\boldsymbol{\mu}, \boldsymbol{x}) = \boldsymbol{x} + \sum_{\boldsymbol{\xi} \in \Xi} \mathbf{c}_{\boldsymbol{\xi}} \Phi_D(\boldsymbol{x}/\eta - \boldsymbol{\xi}), \tag{14}$$

where $\Phi_D(\boldsymbol{x}) : \mathbb{R}^D \rightarrow \mathbb{R}$ represents the cubic D-dimensional B-spline function, $\Xi \subset \mathbb{Z}^D$ is a D-dimensional control-point grid, η is the grid spacing, $\mathbf{c}_{\boldsymbol{\xi}}$ represents the coefficient vector for a control point $\boldsymbol{\xi}$, and the parameter vector $\boldsymbol{\mu}$ is formed by the elements of all coefficient vectors ($\boldsymbol{\mu} = \{\mathbf{c}_{\boldsymbol{\xi}} \mid \boldsymbol{\xi} \in \Xi\}$). Since the B-spline function $\Phi_D(\boldsymbol{x})$ has compact support, the summation effectively goes only over a small neighborhood of control points.

3. Experiments

The experiments were carried out on 2D synthetic Gaussian blob images, 2D cell data, 2D synthetic tagged MRI (tMRI) cardiac image, 3D lung CT and 3D brain MRI scans.

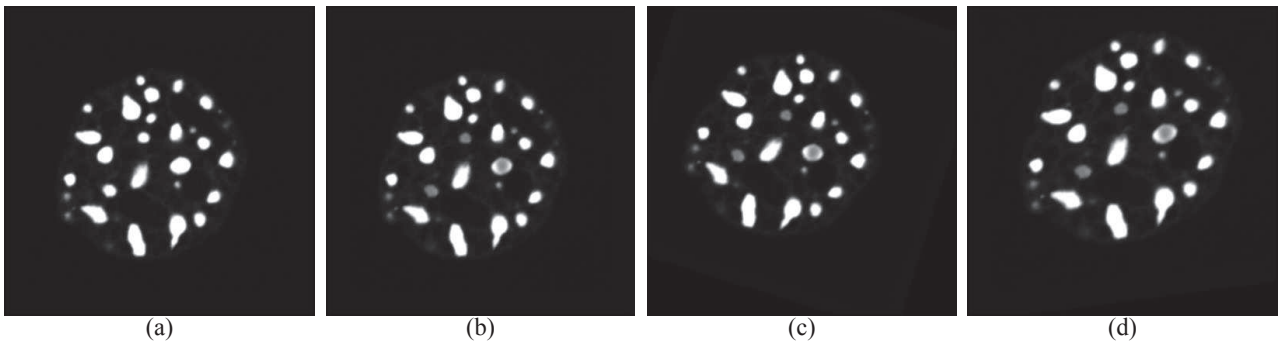

Figure 2: Cell images: (a) frame No.1; (b) frame No.5; (c) frame No.5 with rigid transformation; (d) frame No.5 with affine transformation.

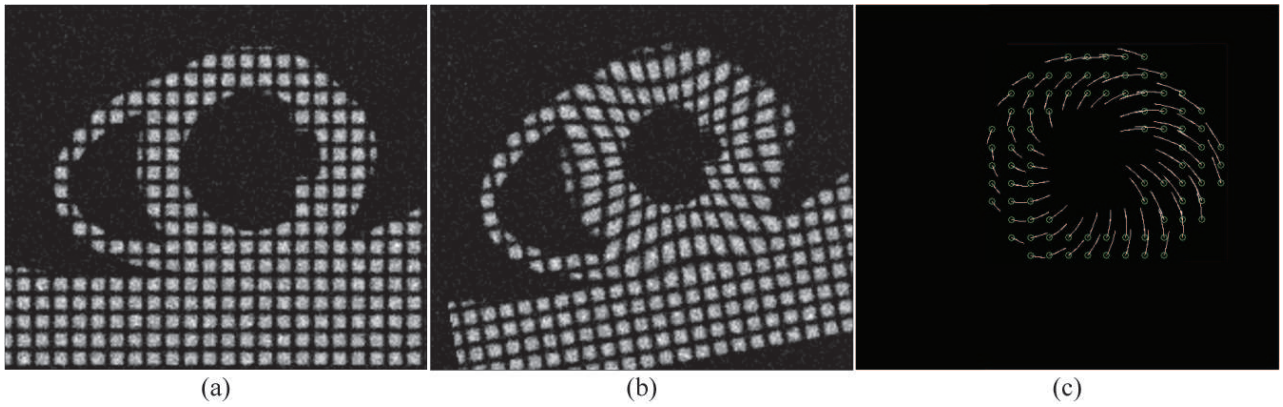

Figure 3: Synthetic tMRI cardiac phantom: (a) frame No.1; (b) frame No.10; (c) ground truth tracks of the tags within the modeled left ventricle.

3.1. Experimental settings

All experiments were implemented based on the open source image registration package `elastix` (Klein et al., 2010). Similarity measure SSD was used as the dissimilarity term on 2D blob images, 2D cell images and 3D lung CT data. On 2D cardiac tMRI images and 3D brain MRI data, NCC was used as the dissimilarity term to handle the image noise and linear intensity changes among scans. Linear interpolation was used to interpolate the moving image. For the ASGD optimizer, the number of random samples S was set to 2000 and 10000 on 2D and 3D data, respectively. The number of iterations K of the optimizer was set to 1000 for all experiments. All experiments with the RS technique were done with numbers of queries: $Q = 1$, $Q = 3$ and $Q = 6$. $\lambda = 0.15$ was used for the exponentially decreasing function h_k. For translation transformation, we added Gaussian noise with $\sigma = 10$mm. For rigid and affine transformations, we applied the RS technique with Gaussian noise $\sigma = 5$mm, $\sigma = 10$mm and $\sigma = 15$mm. For nonrigid B-spline transformation, we added Gaussian noise with $\sigma = \frac{1}{4}\eta$, where η is the control point spacing. The transformation estimated at a coarser scale was used to initialize the transformation on a finer scale when multiresolution strategy is utilized in a experiment. For nonrigid registrations, we repeated all experiments with and without a regularization term that promotes smoothness of the transformation. The commonly used bending energy (Rueckert et al., 1999) was used, with weighting factors of 1 and 10^7 in combination with NCC and SSD, respectively.

3.2. Blob images

Five 2D artificial images with 25 randomly placed Gaussian blobs were created to carry out pair-wise registrations. Figures 1 (a) and (b) give two examples of these images. The synthetic blob images were constructed to simulate a scenario in which the registration cost function exhibits strong local minima. Registration was done directly on the non-blurred original images (i.e., no multiresolution strategy was employed). This allows us to investigate if the RS method can indeed improve the capture range in the presence of strong local minima.

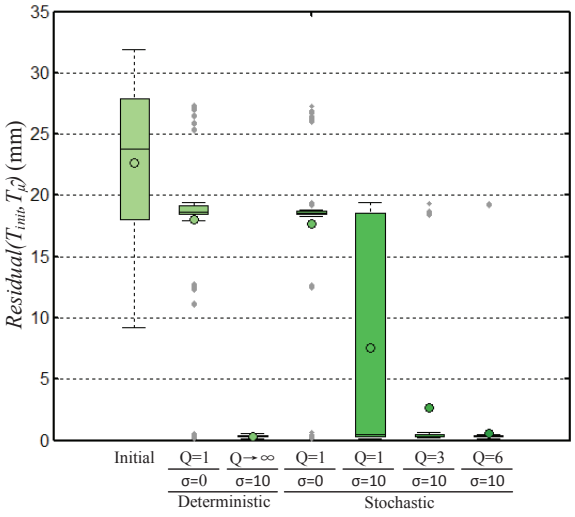

Figure 4: Comparison of the results obtained using deterministic and stochastic approaches on the blob image with translation transform. Boxplots show the distribution of $Residual(\boldsymbol{T}_{init}, \boldsymbol{T}_{\hat{\boldsymbol{\mu}}})$ of 400 test cases using σ = 0mm or 10mm. The whiskers indicate the smallest and largest values within 1.5× the interquartile range and the open circle in each boxplot represents the mean value of all results.

Random translation, rigid, and affine transformations were created and used as initial transformations $\boldsymbol{T}_{init}$. The active transformation $\boldsymbol{T}_{\boldsymbol{\mu}}$ was precomposed with $\boldsymbol{T}_{init}$. The average residual deformation $Residual(\boldsymbol{T}_{init}, \boldsymbol{T}_{\hat{\boldsymbol{\mu}}})$ was defined to measure the registration accuracy by each registration method. The metric measures the average Euclidean distance between the recovered $\boldsymbol{T}_{init}$, i.e., $\boldsymbol{T}_{\hat{\boldsymbol{\mu}}}(\boldsymbol{T}_{init}(\boldsymbol{x}_i))$ and the original location $\boldsymbol{x}_i$:

$$Residual(\boldsymbol{T}_{init}, \boldsymbol{T}_{\hat{\boldsymbol{\mu}}}) = \frac{1}{|\Omega_F|} \sum_{\boldsymbol{x}_i \in \Omega_F} \left\| \boldsymbol{T}_{\hat{\boldsymbol{\mu}}}(\boldsymbol{T}_{init}(\boldsymbol{x}_i)) - \boldsymbol{x}_i \right\|. \tag{15}$$

For 2D translation transformation with $\boldsymbol{\mu} = (t_x, t_y)$, 20 initial transformations were randomly generated within uniform range $t_x, t_y \in [-40, 40]$mm. Therefore, there are in total 4×5×20 test cases for pair-wise registrations on 5 images and 20 initial transformations. For rigid registration with $\boldsymbol{\mu} = (\theta, t_x, t_y)$, 100 initial rigid transformations were randomly generated with uniform range $\theta \in [-0.3, 0.3]$rad, $t_x, t_y \in [-40, 40]$mm. For affine registration with $\boldsymbol{\mu} = (a_{11}, a_{12}, a_{21}, a_{22}, t_x, t_y)$, 100 random affine transformations were generated as the initial transformations. For a_{ij}, random variables were uniformly drawn from range $[-0.2, 0.2]$ and added to an identity matrix. For the translation variables, $t_x, t_y \in [-40, 40]$mm. Therefore, there are in total $4 \times 5 \times 100$ rigid test cases for both pair-wise rigid and affine registrations on 5 images.

To verify whether the proposed stochastic RS technique indeed minimizes Eq. (6), we performed an experiment with a deterministic optimization method, which explicitly tries to minimize Eq. (6) by evaluating the expectation in a brute-force manner. For the deterministic optimization of Eq. (6), all voxels were used (instead of a random subset of S samples in each iteration) to make the optimization deterministic. To numerically approximate the expectation (i.e., integral) in Eq. (6), we exhaustively summed over $\boldsymbol{\Psi}$ in steps of 1mm in x- and y-directions with Gaussian weighting factor (denoted as $Q \to \infty$). Because of the high computational cost of the deterministic validation, we only apply it to the translation transformation.

3.3. Cell images

A series of 2D fluorescence images of a living mouse fibroblast cell were provided by Ghosh et al. (2010). We extracted the first 5 consecutive frames from the 105 images acquired with 2 seconds intervals. The dimensions of these image were resampled to 300×300 pixels. Figures 2 (a) and (b) show frames No.1 and No.5 of these images. Since the first 5 frames are perfectly aligned in the original data, this allows us to evaluate registration accuracy with respect to a ground truth, while using real data. Similar to the experiments on blob images, initial deformations $\boldsymbol{T}_{init}$ were generated and the residual deformation after registration $Residual(\boldsymbol{T}_{init}, \boldsymbol{T}_{\hat{\boldsymbol{\mu}}})$ was computed.

The cell images were registered using a common multiresolution approach, where a Gaussian filter using $\{\sigma_1, \ldots, \sigma_3\} = \{2, 1, 0.5\}$ pixels was adopted to create 3 image resolution levels. Experiments were performed with rigid and affine transformation models. For rigid transformation, 100 initial rigid transformations were randomly generated with uniform range $\theta \in [-0.3, 0.3]$rad, $t_x, t_y \in [-25, 25]$pixels. Pairwise registrations were carried out among these 5 images. For affine transformation, 100 random affine transformations were generated as the initial deformations using $a_{ij} \in [-0.2, 0.2]$ and $t_x, t_y \in [-25, 25]$pixels. Thus, there are in total $4 \times 5 \times 100$ test cases for both rigid and affine registrations. Figures 2 (c) and (d) show examples of the moving image with initial rigid and affine transformation, respectively.

3.4. Synthetic tMRI cardiac images

To assess ventricular function, tMRI is a popular cardiac tissue tracking technique. tMRI enables noninvasive measurement of tissue displacement and deformation of the myocardium by tagging regions of the heart

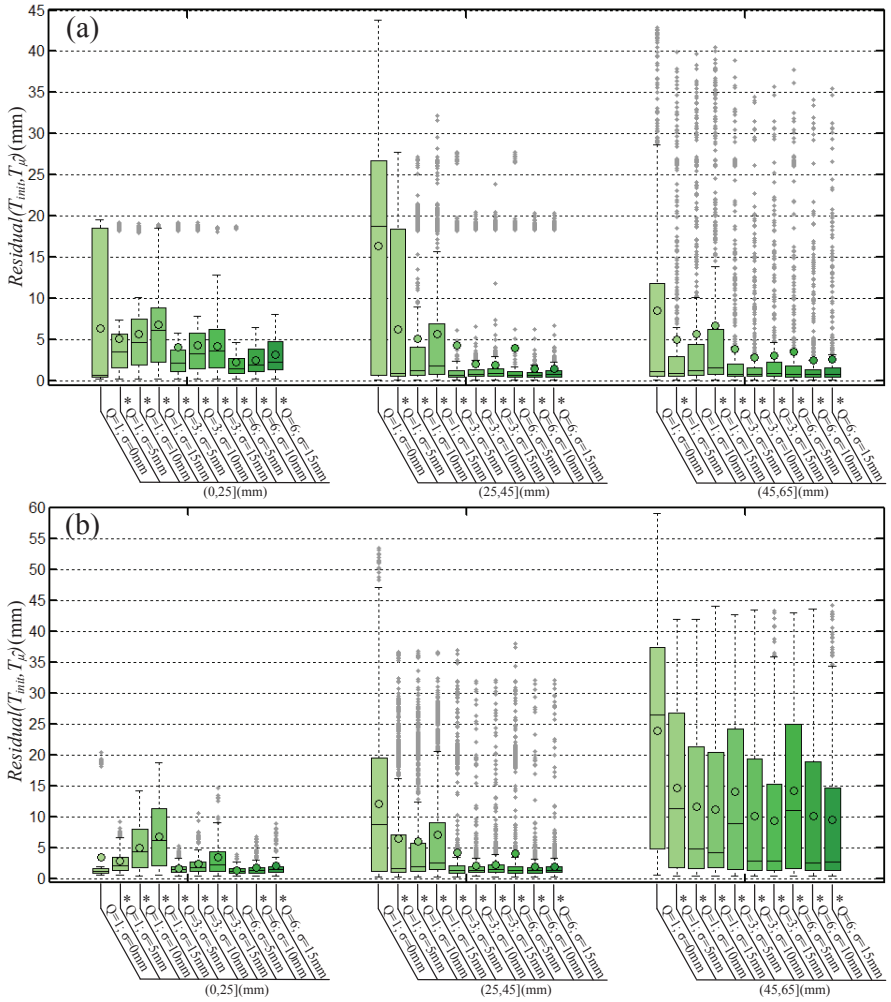

Figure 5: Registration accuracy of rigid and affine registrations on blob images. An asterisk (*) above the label indicates the result is significantly different (paired Wilcoxon signed rank test, $p < 0.05$) from the conventional method ($\sigma = 0$) in the same group. (a) Boxplots of $Residual(\boldsymbol{T}_{init}, \boldsymbol{T}_{\hat{\boldsymbol{\mu}}})$ for rigid registration with each approach; (b) boxplots of $Residual(\boldsymbol{T}_{init}, \boldsymbol{T}_{\hat{\boldsymbol{\mu}}})$ for affine registration with each approach.

wall with two orthogonal sets of magnetic saturation planes, each orthogonal to the image plane. During tissue contraction, the tag pattern moves, allowing visual tracking of the tag-lines over time. Since the deformation of the tag pattern reflects the deformation of the underlying cardiac tissue, robust and accurate localization and tracking of the tags within the images is of clinical significance for assessing dynamic properties of the heart. In this paper we adopted a synthetic phantom (Carranza-Herrezuelo et al., 2010; Smal et al., 2012) to evaluate the RS method in a controlled, but fairly realistic setting, involving nonrigid deformations. Figures 3 (a) and (b) show two examples, one without deformation and one with the largest deformation among the 30 frames of the simulated cardiac cycle. On each frame, Rician noise was added in order to create images with the same signal-to-noise ratio of 18 dB. Figure 3 (c) plots the ground truth tracks of the tags of the modeled left ventricle. It can be found that the nonrigid deformation contains radial expansion, contraction and rotation to simulate left ventricular contraction through the cardiac cycle. Note that the tagging lines may cause local minima during image registration. We did pairwise registrations among all 30 frames instead of only between neighboring frames. In total 29×30 nonrigid registration experiments were thus performed. A Gaussian filter using $\{\sigma_1, \ldots, \sigma_3\} = \{2, 1, 0.5\}$ pixels was applied to create 3 image resolution levels. Also for the B-spline transformation model, we employed a standard multiresolution scheme, halving the grid spacing η in

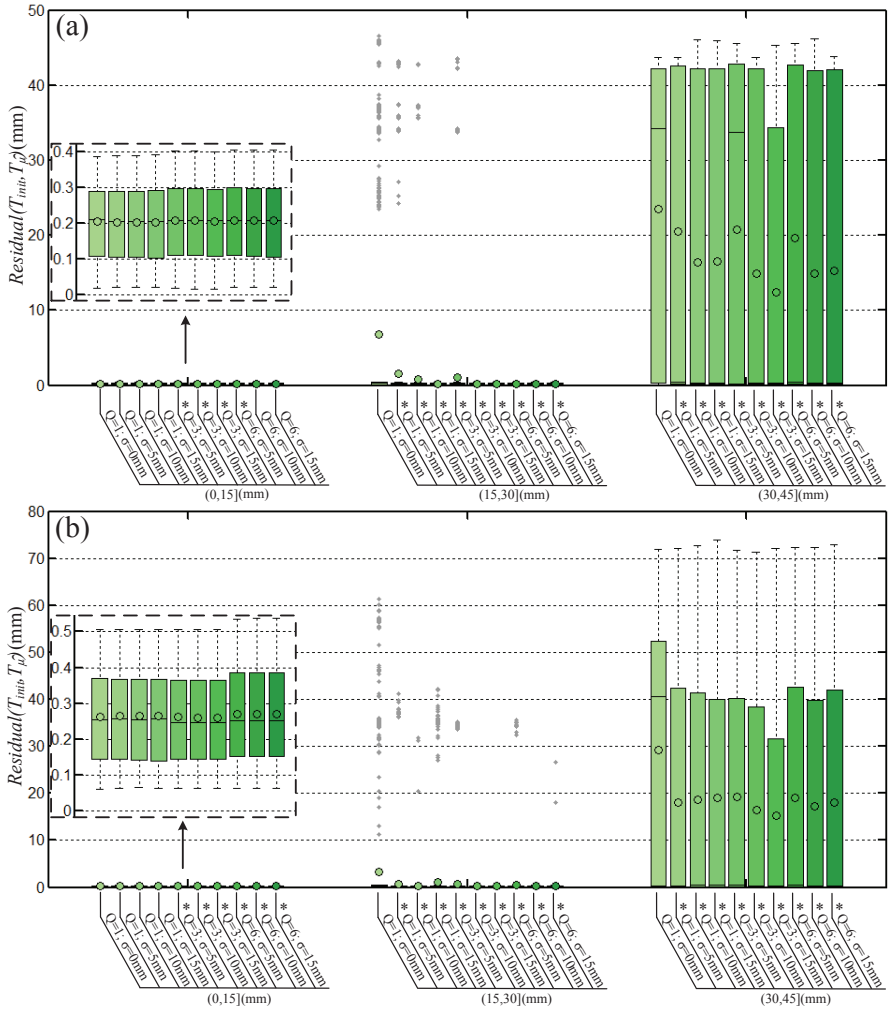

Figure 6: Registration accuracy of rigid and affine registrations on cell images. An asterisk (*) above the label indicates the result is significantly different (paired Wilcoxon signed rank test, $p < 0.05$) from the conventional method ($\sigma = 0$) in the same group. (a) Boxplots of $Residual(\boldsymbol{T}_{init}, \boldsymbol{T}_{\hat{\boldsymbol{\mu}}})$ for rigid registration with each approach; (b) boxplots of $Residual(\boldsymbol{T}_{init}, \boldsymbol{T}_{\hat{\boldsymbol{\mu}}})$ for affine registration with each approach.

each resolution level: $\{\eta_1, \eta_2, \eta_3\} = \{64, 32, 16\}$mm. The mean of target registration errors (TRE), which measure the distances between the transformed and ground truth landmarks, was used to measure the registration accuracy. The centers of tag-line intersections (Smal et al., 2012) were used as landmarks.

3.5. Lung CT images

Public DIR-Lab 3D chest CT data enables a rigorous and objective assessment of the spatial accuracy of registration methods (Castillo et al., 2009). The DIR-Lab data set contains 10 pairs of scans with 300 manually annotated landmarks on the lungs, which allows us to evaluate the registration accuracy obtained by RS on real data. The voxel sizes and dimensions of these scans are around 1.0×1.0×2.5mm and around 256×256×110 voxels. To focus on the lung region, lung masks were created to restrict the registration. The masks were created by thesholding, 3D-6-neighborhood connected component analysis, and morphological closing operation using a spherical kernel with a diameter of 9 voxels. In the experiments, the exhale phase (moving image) was registered to the inhale phases (fixed image). A Gaussian filter using $\{\sigma_1, \ldots, \sigma_4\} = \{4, 2, 1, 0.5\}$ voxels was used to create 4 image resolution levels.

Registration experiments on lung CT images were done with a B-spline transformation. We fixed the coarsest grid spacing η_1 to be 64mm to avoid too large grid spacing, and we used η_4 = 8mm and η_4 = 10mm as the finest grid spacings. From η_1 and η_4, the

Table 1: Registration robustness using blob images for rigid and affine registrations, measured by the percentage of cases where $Residual(\boldsymbol{T}_{init}, \boldsymbol{T}_{\hat{\mu}}) \leq 2$mm.

		$Q = 1$	$Q = 1$			$Q = 3$			$Q = 6$		
	Initial $Residual(\boldsymbol{T}_{init}, \boldsymbol{T}_{\hat{\mu}})$	$\sigma = 0mm$	$5mm$	$10mm$	$15mm$	$5mm$	$10mm$	$15mm$	$5mm$	$10mm$	$15mm$
Rigid	(0, 25]mm	**68%**	33%	25%	19%	47%	38%	35%	63%	53%	45%
	(25, 45]mm	41%	69%	64%	53%	80%	87%	83%	82%	**94%**	92%
	(45, 65]mm	60%	69%	63%	56%	74%	78%	74%	76%	**81%**	78%
	Overall	51%	63%	58%	48%	73%	77%	73%	77%	**84%**	81%
Affine	(0, 25]mm	87%	41%	27%	24%	75%	57%	46%	**91%**	78%	74%
	(25, 45]mm	41%	57%	50%	39%	73%	77%	67%	75%	**81%**	77%
	(45, 65]mm	19%	29%	34%	30%	34%	41%	39%	32%	**45%**	42%
	Overall	39%	47%	43%	35%	61%	64%	56%	64%	**70%**	66%

Table 2: Registration robustness using cell images for rigid and affine registrations, measured by the percentage of cases where $Residual(\boldsymbol{T}_{init}, \boldsymbol{T}_{\hat{\mu}}) \leq$ 2mm.

		$Q = 1$	$Q = 1$			$Q = 3$			$Q = 6$		
	Initial $Residual(\boldsymbol{T}_{init}, \boldsymbol{T}_{\hat{\mu}})$	$\sigma = 0mm$	$5mm$	$10mm$	$15mm$	$5mm$	$10mm$	$15mm$	$5mm$	$10mm$	$15mm$
Rigid	(0, 15]mm	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
	(15, 30]mm	80%	96%	98%	**100%**	98%	**100%**	**100%**	**100%**	**100%**	**100%**
	(30, 45]mm	45%	50%	50%	52%	49%	64%	64%	59%	**70%**	**70%**
	Overall	60%	69%	71%	72%	71%	77%	76%	75%	**80%**	**80%**
Affine	(0, 15]mm	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
	(15, 30]mm	92%	99%	99%	99%	**100%**	**100%**	**100%**	97%	**100%**	**100%**
	(30, 45]mm	40%	63%	59%	57%	58%	64%	62%	65%	**68%**	63%
	Overall	64%	72%	69%	67%	73%	74%	71%	74%	**77%**	74%

multiresolution schedule was calculated according to $\{\eta_1, \eta_4(\eta_1/\eta_4)^{2/3}, \eta_4(\eta_1/\eta_4)^{1/3}, \eta_4\}$. Therefore, the grid schedules were set to $\{\eta_1, \eta_2, \eta_3, \eta_4\}$=$\{64, 34, 19, 10\}$mm and $\{\eta_1, \eta_2, \eta_3, \eta_4\}$=$\{64, 32, 16, 8\}$mm from coarsest to the finest grid spacings. The two slightly different schedules were used to verify the consistency of the (relative) performances of different registration methods. TRE of the landmarks was used to measure the registration accuracy. In addition to accuracy, the standard deviation of the determinant of the spatial Jacobian (D_{SJ}) was utilized to evaluate the transformation smoothness. D_{SJ} represents the local volume change at a specific location. Its standard deviation measures the variation of local compression and expansion, and thus provides an indication of the smoothness of a transformation.

3.6. Brain MRI images

Intersubject brain registration is a challenging task in medical image analysis. It is widely used for atlas based segmentation (Heckemann et al., 2006) and template construction (Mandal et al., 2012). We used the public Internet brain segmentation repository (IBSR v2.0), which contains 18 T1-weighted MRI 3D brain scans, to evaluate the RS method in the setting of intersubject registration. The volumes of these images are $256 \times 256 \times 128$ voxels. The voxel sizes are around $1 \times 1 \times 1.5$mm. A Gaussian filter using $\{\sigma_1, \ldots, \sigma_4\} = \{4, 2, 1, 0.5\}$ voxels was used to create 4 image resolution levels. The same affine registrations were used to roughly align the brain data, and then these initialized results were used as input data for the B-spline registrations. All experiments were repeated with grid schedules $\{40, 20, 10, 5\}$mm and $\{24, 12, 6, 3\}$mm, to verify the consistency of the (relative) performances of different registration methods. To evaluate the registration accuracy, overall mean overlap which measures the overlap between the transformed and ground truth atlases over all labels was used (Klein et al., 2009a). We used the standard deviation of D_{SJ} to evaluate the smoothness of the transformation inside the brain mask. Because intersubject registration on 18 patients was performed there were in total 306 test cases.

4. Results

4.1. Blob images

Figure 4 shows the experimental results on blob images using translation transformation. It shows the ini-

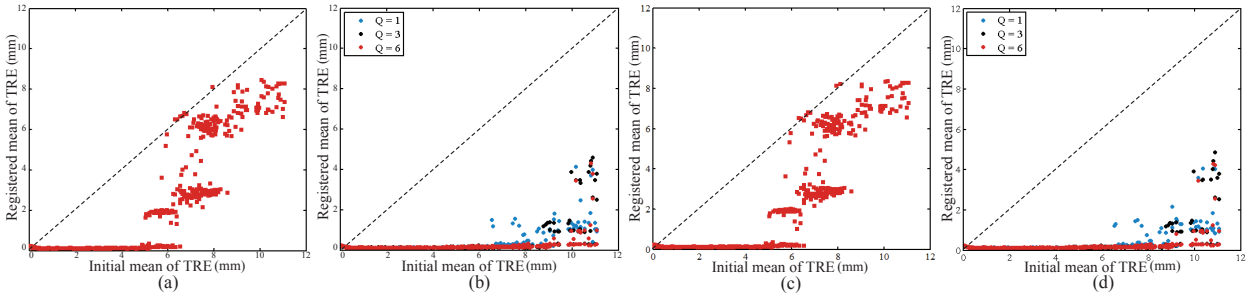

Figure 7: Registration accuracy on synthetic tMRI cardiac images: (a) and (c) registration accuracy by the conventional method ($\sigma = 0$), without and with regularization term respectively; (b) and (d) registration accuracy produced by the RS technique ($\sigma = \frac{1}{4}\eta$) using $Q = 1$, 3 and 6, without and with regularization term respectively. Each dot represents a test case based on a pair of images.

Table 3: TREs (in mm, mean ± st.dev.) achieved by different unregularized approaches on 10 lung CT pairs. Results for B-spline grid spacings $\eta_4 = 10$mm and $\eta_4 = 8$mm are shown. The best results in each row for each η_4 are marked bold.

		$\eta_4 = 10$mm				$\eta_4 = 8$mm			
No.	Initial	σ=0	Q=1	Q=3	Q=6	σ=0	Q=1	Q=3	Q=6
1	3.89±2.8	**0.98±0.5**	**0.98±0.5**	**0.98±0.5**	**0.98±0.5**	**0.99±0.5**	**0.99±0.5**	**0.99±0.5**	**0.99±0.5**
2	4.34±3.9	**0.96±0.5**	**0.96±0.5**	**0.96±0.5**	**0.96±0.5**	**0.96±0.5**	0.97±0.5	0.98±0.5	0.98±0.5
3	6.94±4.1	1.13±0.6	**1.12±0.6**	1.13±0.6	**1.12±0.6**	**1.14±0.6**	1.16±0.6	1.15±0.6	**1.14±0.6**
4	9.83±4.9	2.98±4.4	**2.36±3.0**	2.40±3.2	2.45±3.3	3.53±5.1	2.99±4.1	**2.93±4.2**	2.94±4.2
5	7.48±5.5	1.38±1.3	**1.35±1.2**	**1.35±1.2**	**1.35±1.2**	1.37±1.3	**1.34±1.2**	**1.34±1.2**	**1.34±1.2**
6	10.89±7.0	3.79±5.5	3.62±5.2	3.65±5.2	**3.62±5.1**	4.43±5.9	**4.17±5.6**	4.23±5.6	4.44±5.8
7	11.03±7.4	1.55±1.2	**1.49±1.0**	1.50±1.0	1.50±1.0	1.71±1.6	1.64±1.4	**1.60±1.3**	1.64±1.4
8	14.99±9.0	2.17±3.8	2.19±3.8	2.14±3.7	**2.08±3.5**	2.29±4.0	2.35±4.0	2.29±3.9	**2.24±3.8**
9	7.92±4.0	1.81±1.9	1.61±1.5	**1.58±1.4**	1.71±1.8	2.31±2.9	**2.04±2.4**	2.12±2.6	2.10±2.6
10	7.30±6.4	1.32±0.9	**1.31±0.9**	**1.31±0.9**	**1.31±0.9**	1.32±1.0	**1.32±0.9**	**1.32±0.9**	1.33±0.9
Mean	8.46±3.3	1.81±0.9	**1.70±0.8**	**1.70±0.8**	1.71±0.8	2.01±1.2	**1.90±1.0**	**1.90±1.0**	1.91±1.1

tial $Residual(\boldsymbol{T}_{init}, \boldsymbol{T}_{\hat{\boldsymbol{\mu}}})$, the results of the deterministic approaches (see Section 3.2), and the results of the stochastic methods with and without RS on translation transformations. For the deterministic optimization, both the registration results achieved by the ordinary deterministic ($Q = 1$; $\sigma = 0$) and the brute-force ($Q \to \infty$; $\sigma = 10$) methods are presented. We can find the ordinary method obtained unsatisfactory results with large $Residual(\boldsymbol{T}_{init}, \boldsymbol{T}_{\hat{\boldsymbol{\mu}}})$. However, the brute-force approach, which directly minimizes the smoothed cost function (Eq. (6)), reached the global optimum with almost zero $Residual(\boldsymbol{T}_{init}, \boldsymbol{T}_{\hat{\boldsymbol{\mu}}})$. For the stochastic group, the ordinary stochastic approach ($Q = 1$; $\sigma = 0$) obtained similar results as the ordinary deterministic method. The proposed RS method improved the registration results and, when $Q = 3$ or $Q = 6$, achieved similar registration accuracy as the costly brute-force manner.

Figures 5 (a) and (b) show the results by different registration approaches on rigid and affine transformations. The results are grouped by the magnitude of the initial transformation (i.e., $Residual(\boldsymbol{T}_{init}, \boldsymbol{T}_{\hat{\boldsymbol{\mu}}})$ before registration). In most cases, the results with RS ($\sigma > 0$) achieved better accuracy than the conventional method ($\sigma = 0$). For small initial transformations, RS lead to somewhat worse median accuracies, especially with high σ and low Q. Paired Wilcoxon signed rank tests (Hollander and Wolfe, 1999) were performed to verify the statistical significance of the registration results. Significance is indicated by an asterisk in this paper. The paired comparisons were carried out between the conventional method ($\sigma = 0$) and the methods with the RS technique ($\sigma > 0$) inside each group. It was found that the methods with the RS technique produced significantly better results than the conventional method in the cases with medium and large initial transformations. Among the methods with the RS technique, registration accuracy was usually further enhanced by increasing the number of queries Q.

To summarize the robustness of different registration methods, Tables 1 provides an overview of the percentages of the results with $Residual(\boldsymbol{T}_{init}, \boldsymbol{T}_{\hat{\boldsymbol{\mu}}}) \leq 2$mm using rigid and affine transformations, respectively. In most

Table 4: TREs (in mm, mean ± st.dev.) achieved by different regularized approaches on 10 lung CT pairs. Results for B-spline grid spacings η_4 = 10mm and η_4 = 8mm are shown. The best results in each row for each η_4 are marked bold.

		η_4 = 10mm				η_4 = 8mm			
No.	Initial	σ=0	Q=1	Q=3	Q=6	σ=0	Q=1	Q=3	Q=6
1	3.89±2.8	**0.97±0.5**	**0.97±0.5**	**0.97±0.5**	**0.97±0.5**	**0.97±0.5**	**0.97±0.5**	**0.97±0.5**	**0.97±0.5**
2	4.34±3.9	**0.94±0.5**	**0.94±0.5**	**0.94±0.5**	0.95±0.5	**0.94±0.5**	0.95±0.5	0.95±0.5	0.95±0.5
3	6.94±4.1	**1.10±0.6**	**1.10±0.6**	**1.10±0.6**	1.12±0.6	1.11±0.6	**1.10±0.6**	**1.10±0.6**	**1.10±0.6**
4	9.83±4.9	2.32±3.0	**1.99±2.2**	2.01±2.3	2.02±2.3	2.73±3.9	**2.21±2.8**	2.29±3.0	2.29±3.0
5	7.48±5.5	1.37±1.2	1.35±1.2	**1.34±1.2**	1.35±1.2	1.34±1.2	**1.32±1.2**	**1.32±1.2**	1.33±1.2
6	10.89±7.0	3.33±4.8	**3.02±4.3**	3.18±4.5	3.07±4.4	3.71±5.2	**3.60±5.1**	3.64±5.1	3.70±5.3
7	11.03±7.4	1.48±1.1	**1.43±1.0**	**1.43±1.0**	1.44±1.0	1.55±1.3	1.46±1.0	**1.45±1.0**	1.48±1.1
8	14.99±9.0	2.12±3.7	2.16±3.7	2.12±3.6	**2.03±3.4**	2.21±3.9	2.28±3.9	2.22±3.8	**2.17±3.7**
9	7.92±4.0	1.45±1.1	**1.36±0.9**	**1.36±0.9**	1.40±1.1	1.74±1.7	1.59±1.5	1.61±1.6	**1.55±1.4**
10	7.30±6.4	1.29±0.9	1.28±0.9	1.28±0.9	**1.28±0.8**	1.27±0.9	1.27±0.9	**1.27±0.8**	1.27±0.9
Mean	8.46±3.3	1.64±0.7	**1.36±0.8**	1.57±0.7	1.56±0.7	1.76±0.9	**1.67±0.8**	1.68±0.8	1.68±0.8

Table 5: Transformation smoothness obtained by different unregularized registration approaches on 10 lung CT pairs. Results for B-spline grid spacings η_4 = 10mm and η_4 = 8mm are shown. The best results in each row for each η_4 are marked bold.

	η_4 = 10mm				η_4 = 8mm			
No.	σ=0	Q=1	Q=3	Q=6	σ=0	Q=1	Q=3	Q=6
1	**0.10**	**0.10**	**0.10**	**0.10**	**0.12**	0.13	0.13	**0.12**
2	**0.10**	**0.10**	**0.10**	**0.10**	**0.12**	**0.12**	**0.12**	**0.12**
3	**0.12**	**0.12**	**0.12**	**0.12**	**0.14**	0.15	0.15	0.15
4	0.31	**0.25**	0.26	0.27	0.40	0.38	**0.37**	0.38
5	0.16	**0.15**	0.16	**0.15**	**0.18**	0.19	**0.18**	**0.18**
6	**0.36**	0.37	**0.36**	**0.36**	0.48	**0.45**	0.47	0.48
7	0.20	**0.19**	**0.19**	0.20	**0.25**	**0.25**	**0.25**	**0.25**
8	**0.18**	**0.18**	**0.18**	**0.18**	**0.21**	0.22	0.22	0.22
9	0.21	**0.19**	**0.19**	0.20	0.32	**0.29**	0.31	0.30
10	**0.17**	**0.17**	**0.17**	**0.17**	**0.20**	0.21	0.21	0.21
Mean	0.19±0.1	**0.18±0.1**	**0.18±0.1**	0.19±0.1	**0.24±0.1**	**0.24±0.1**	**0.24±0.1**	**0.24±0.1**

cases, the RS technique enhanced the registration robustness, except for small initial transformations. Overall, the RS approach with $\sigma = 10$ and $Q = 6$ achieved the best robustness.

4.2. Cell images

Figures 6 (a) and (b) show the registration results by different registration approaches using rigid and affine transformations on cell data. Like in Figure 5, the results are grouped by initial degree of transformation before registration. It can be observed that the approaches with the RS technique improve the registration results in most cases. For cases with small initial transformation magnitude, the accuracy of the original method was maintained. The paired Wilcoxon signed rank tests were carried out between the conventional method ($\sigma = 0$) and the methods with the RS technique ($\sigma > 0$) inside each group.

The registration results with $Residual(\boldsymbol{T}_{init}, \boldsymbol{T}_{\hat{\mu}}) \leq$ 2mm are summarized in Tables 2 for rigid and affine transformations, respectively. For the cases with small initial transformation, all methods can register well. For the cases with medium and large initial transformations, the approaches with the RS technique consistently achieved better registration robustness than the conventional method. Overall, the RS approach with $\sigma = 10$ and $Q = 6$ achieved the best robustness.

4.3. Synthetic tMRI cardiac images

Figure 7 show the registration results achieved by the conventional method ($\sigma = 0$) and the methods with the RS technique ($\sigma = \frac{1}{4}\eta$) using different values of Q. Figures 7 (a) and (b) presents the results without regularization term, whereas Figures 7 (c) and (d) show the results obtained by adding a regularization term to the cost function. The proposed RS technique obtained much better registration accuracy than the conventional

Table 6: Transformation smoothness obtained by different regularized registration approaches on 10 lung CT pairs. Results for B-spline grid spacings η_4 = 10mm and η_4 = 8mm are shown. The best results in each row for each η_4 are marked bold.

	η_4 = 10mm				η_4 = 8mm			
No.	σ=0	Q=1	Q=3	Q=6	σ=0	Q=1	Q=3	Q=6
1	**0.09**	**0.09**	**0.09**	**0.09**	**0.09**	**0.09**	**0.09**	**0.09**
2	**0.08**	**0.08**	**0.08**	**0.08**	**0.08**	**0.08**	**0.08**	**0.08**
3	**0.10**	**0.10**	**0.10**	**0.10**	**0.10**	**0.10**	**0.10**	**0.10**
4	0.20	**0.18**	0.19	0.19	0.24	**0.21**	**0.21**	**0.21**
5	**0.13**	**0.13**	**0.13**	**0.13**	**0.14**	**0.14**	**0.14**	**0.14**
6	0.29	**0.27**	0.29	0.28	0.35	**0.32**	**0.32**	0.33
7	**0.17**	**0.17**	**0.17**	**0.17**	**0.19**	**0.19**	**0.19**	**0.19**
8	**0.16**	**0.16**	**0.16**	**0.16**	**0.17**	0.18	0.18	0.18
9	**0.15**	**0.15**	**0.15**	**0.15**	0.18	**0.17**	0.18	**0.17**
10	**0.15**	**0.15**	**0.15**	**0.15**	**0.17**	**0.17**	**0.17**	**0.17**
Mean	**0.15±0.1**	**0.15±0.1**	**0.15±0.1**	**0.15±0.1**	**0.17±0.1**	**0.17±0.1**	**0.17±0.1**	**0.17±0.1**

one in both cases with or without regularization terms. It should be noted that the regularization term did not help to avoid the local minima in Figure 7 (c). For the test cases with small initial means of TREs (small values along X-axis), the final registration accuracy was not distracted by the added noise. Meanwhile, the registration results for large initial means of TREs were substantially improved by using the RS technique.

4.4. Lung CT images

Table 3 presents the registration accuracy TRE achieved by registration methods with different values of σ and Q without using a regularization term. Registration methods with the RS technique ($\sigma = \frac{1}{4}\eta$) achieved slightly better overall TREs than the conventional method ($\sigma = 0$). Even though the differences of accuracy are small, these improvements are consistently found for both η_4 = 10mm and η_4 = 8mm. The results of registration accuracy with the regularization term are shown in Table 4. The overall results were improved by regularizing the registration process. Again, the mean results achieved with the RS technique were consistently better than the conventional method ($\sigma = 0$). Compared with a recent study (Papież et al., 2014) on the same data, the TREs reported here are competitive. The registration accuracy was not further enhanced by increasing Q. Table 5 reports the transformation smoothness obtained by each registration approach. The RS method produced similar results as the conventional approach. Because the regularization term promotes smooth transformations, the transformation smoothness was improved significantly as shown in Table 6. In these results the RS approaches still obtained similar smoothness as the conventional method.

4.5. Brain MRI images

Figure 8 shows the registration accuracy (Figure 8 (a)) and transformation smoothness (Figure 8 (b)) by different approaches without and with regularization term. Since we have sufficient (306) test cases in this experiment, paired Wilcoxon signed rank tests were performed to verify the statistical significance of the registration results. The paired comparisons were carried out between the traditional approach ($\sigma = 0$) and the methods with the RS technique. Figure 8 (a) shows that the registration accuracies achieved with the RS method are statistically significantly better than the accuracies without the RS method. Even though these differences are small in magnitude (around 1% in overlap), the improvements of registration accuracy are consistently found for both η_4 = 5mm and η_4 = 3mm. The overall registration accuracy was not improved in the regularized experiments. Figure 8 (b) shows that the results of transformation smoothness are practically identical without and with RS technique, despite being statistically different. As expected, adding the regularization term improved transformation smoothness.

5. Discussion

We evaluated the feasibility of using the stochastic RS technique in image registration. The RS technique was applied to various registration problems including translation, rigid, affine and B-spline transformation models. Extensive experiments were carried out on both synthetic and real biomedical data. The improvement in registration results proves the effectiveness of the RS technique. In theory the RS technique approximates a convolution of the cost function with a Gaussian PDF.

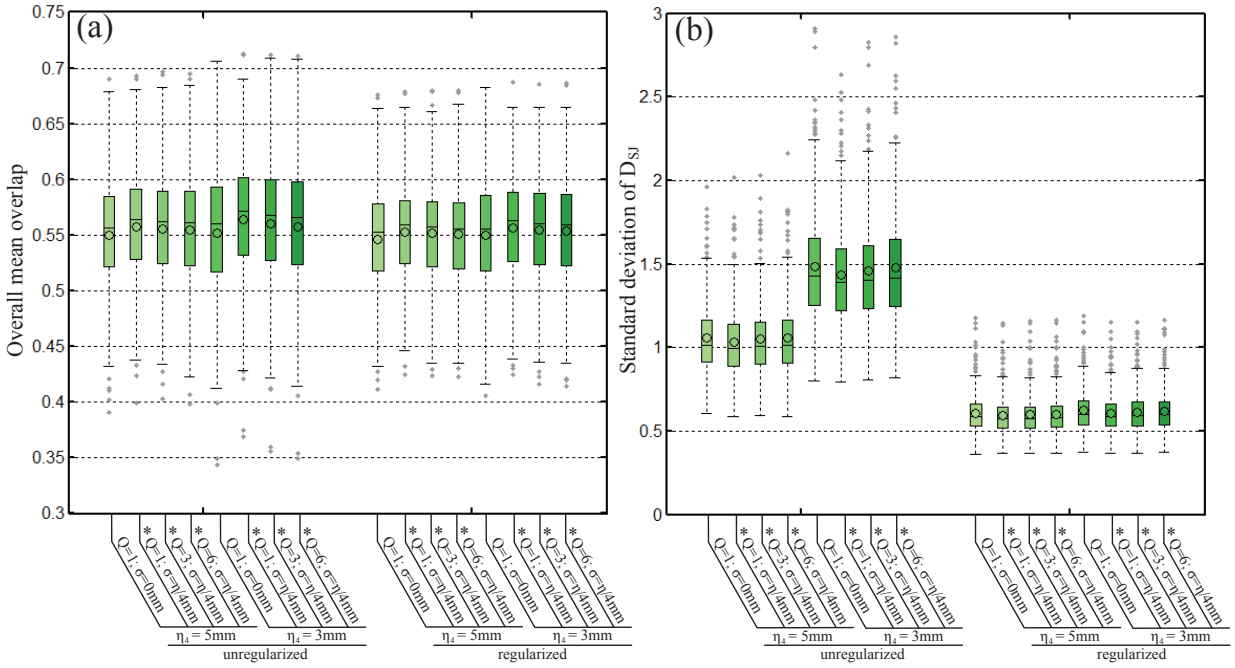

Figure 8: Registration results on brain MRI data when using different values of σ and Q. An asterisk (*) below the label indicates a significant difference (paired Wilcoxon signed rank test, $p < 0.05$) from $\sigma = 0$mm. (a) accuracy of unregularized and regularized registration (overall mean overlap, higher values are better); (b) transformation smoothness of unregularized and regularized registration (standard deviation of D_{SJ}, lower values are better). Results for B-spline grid spacings $\eta_4 = 5$mm and $\eta_4 = 3$mm are shown.

Thus, the technique helps to avoid local minima while minimizing the cost function of the image registration.

The brute-force approach discussed in Section 3.2 was implemented as a deterministic reference method, which directly smoothes the cost function by convolving it with a Gaussian PDF, according to Eq. (6). The registration results obtained using this deterministic method confirm that improved smoothness of the cost function leads to better registration accuracy. The proposed stochastic RS method obtained very similar registration accuracy as the deterministic reference method. This indicates that the RS method has an equivalent cost function smoothing effect, as predicted by theory. The brute-force manner was feasible for translations in 2D, but becomes computationally impractical for high-dimensional transformation models because of the high computational cost. In contrast, the RS approach is an efficient way to minimize Eq. (6) in high-dimensional settings.

This is the first time that the RS technique is introduced in the image registration domain, where local minima often lead to misregistrations. To apply the technique in this domain, we made the following contributions to the original RS technique (Duchi et al., 2012): 1) the rotation, scale and shear parameters have totally different ranges from the translation parameters in rigid and affine transformations; therefore, we applied an automatically estimated diagonal scaling matrix to normalize the optimization process; 2) we employed an exponentially decreasing function to control the amplitude of Gaussian noise during optimization; 3) the RS technique is combined with the ASGD optimizer (Klein et al., 2009b), which is a widely used stochastic optimizer in image registration. Finally, exhaustive experiments were performed on both synthetic and real images, using translation, rigid, affine and B-spline transformations, both with and without regularization term, in order to provide detailed insight into the value of RS for image registration.

Except for the experiments on artificial blob images, a standard multiresolution strategy was incorporated in all experiments. As an established and well-proven technique, a multiresolution strategy is usually considered an essential component in image registration. However, it cannot remove all the local minima in the optimization landscape of image registration (Jenkinson and Smith, 2001). The experimental results show that the proposed RS technique is complementary to the multiresolution strategy for further reducing the local minima in image registration. As a

generic approach, RS technique is not only constrained for intensity-based registrations but also for more generalized registration tasks, such as point-based registrations (Besl and McKay, 1992; Chui and Rangarajan, 2003; Jian and Vemuri, 2011).

In case of larger initial transformations, the chance to encounter multiple local minima due to image structures with similar appearance increases. The results on blob, cell and tMRI data clearly show that the RS technique increased the success rates for the cases with medium and large initial transformations. A 100% success rate was not achieved, since for very large-scale local minima it could be impossible to overcome them. Intuitively, the size of local minima that can be overcome is proportional to the setting of σ, which controls the magnitude of random perturbations. In practice, the value of σ cannot be increased indefinitely, since the optimization process may start to diverge.

In contrast to the traditional regularization term that is often added to the cost function in nonlinear registration problems, the proposed RS technique provides a direct way to tackle local minima encountered in both linear and nonlinear registration problems. In the experiments on synthetic tMRI cardiac images, the bending energy regularization term did not help to avoid the local minima, whereas the RS technique improved the accuracy substantially, both with and without the regularization term. On the lung CT images, regularization led to improved accuracy and transformation smoothness, but the RS technique led to even further improvement on that result in terms of accuracy. On brain MRI data, regularization improved transformation smoothness, but did not improve accuracy. RS resulted in a slight improvement in accuracy, both with and without regularization. From these results, it becomes clear that the regularization term and the RS technique target different aspects.

For the blob images, the registration accuracy obtained by the RS technique deteriorated in some cases with small initial transformations and higher values of σ (see Figure 5). However, this was not observed for the cell images and other data types. A possible explanation for this unsatisfactory result is that the large Gaussian noise prohibits optimal convergence. One possible solution is to finetune the exponentially decreasing function h_k, e.g., using a smaller λ to enforce faster decay. In this paper, however, we kept h_k consistent for all experiments.

In most experiments with medium to large initial deformations, increasing Q improved the accuracy. In a few experiments, however, increasing Q seems to slightly deteriorate the accuracy. This could be seen in Figure 6 for cell images with small initial deformations, in Tables 3 and 4 for lung CT data, and in Figure 8 for brain MRI. These unexpected results might be related to the automatically estimated step sizes a_k used by the ASGD optimizer. The optimal step size depends on the amount of noise on the stochastic gradient. For high Q, the variance of $\tilde{\boldsymbol{g}}$ is reduced, and therefore a higher step size might be optimal. This is a possible direction for future research.

Compared with the other presented applications, the relatively small improvement on CT lung and MRI brain data could be explained by less number of local minima presented in the cost function. The multiresolution strategy could already effectively avoid many of them. Despite this, consistent and statistically significant small improvements were observed in the experiments.

Regarding computational cost, all experiments with the RS technique were done with query numbers $Q = 1$, 3 and 6. For the method with $Q = 1$, the computational cost is kept the same as for the traditional method. For the method with multiple queries ($Q = 3$ and $Q = 6$) the computation cost is increased by a factor Q. However, the RS technique itself is very suitable to parallel computation since multiple queries are independent of each other. The experimental results indicate that for not too large σ, $Q = 1$ is sufficient.

In our previous studies (Sun et al., 2014b,a) a perturbation technique was derived from the convolutional relationship of B-splines, in order to enable the use of computationally efficient lower-order B-spline basis functions for nonrigid image registration. In these works the entire B-spline grid is shifted to approximate a higher-order B-spline transformation. However, the transformation parameters themselves were not perturbed.

The random coordinate resampling techniques (Likar and Pernuš, 2001; Tsao, 2003; Aganj et al., 2013) are, in a way, based on a similar principle as the RS method, as they effectively smooth the cost function by perturbing the image coordinates. The RS technique presented in our work is more generically applicable, since it not only focuses on small-scale local minima due to interpolation artifacts, but also on large-scale local minima, caused for example by image content.

In future work we plan to apply the RS technique to more registration tasks. The RS technique treats the cost function as a black box; it works directly on the parameters $\boldsymbol{\mu}$. Therefore, in principle the RS technique can be used in combination with any transformation model and similarity measure. Whereas in this work, we restricted ourselves to intensity-based registration methods, it could be an interesting future research direction to apply the method to point-based registration methods.

For instance, a general framework for point set registration is presented and incorporated with rigid and non-rigid transformation models by Jian and Vemuri (2011). In their work, it was observed that the L2 distance for measuring similarity between two Gaussian mixtures may cause local minima in the optimization process. Thus, it could be interesting to apply the RS technique to these types of point set registration problems as well. In addition, developing a mechanism to automatically determine the optimal amplitude of added noise could be an interesting direction.

6. Conclusion

In this work we developed a RS technique for image registration. This technique consists of adding random noise to the transformation parameters during the stochastic optimization of the registration parameters. It is shown that the RS technique smoothes the cost function, thus suppressing local minima and thereby improving the chance of finding the global minimum. By gradually decreasing the noise level during optimization, the local optimization performance is maintained. The improved registration results demonstrate the effectiveness of the RS technique.

7. Acknowledgements

Wei Sun would like to acknowledge funding from National Natural Science Foundation of China (U1301251).

References

Aganj, I., Yeo, B.T., Sabuncu, M.R., Fischl, B., 2013. On removing interpolation and resampling artifacts in rigid image registration. IEEE Transactions on Image Processing 22, 816–827.

Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Medical Image Analysis 12, 26–41.

Beck, A., Teboulle, M., 2012. Smoothing and first order methods: A unified framework. SIAM Journal on Optimization 22, 557–580.

Besl, P., McKay, N.D., 1992. A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence 14, 239–256.

Bhagalia, R., Fessler, J., Kim, B., et al., 2009. Accelerated nonrigid intensity-based image registration using importance sampling. IEEE Transactions on Medical Imaging 28, 1208–1216.

Carranza-Herrezuelo, N., Bajo, A., Sroubek, F., Santamarta, C., Cristóbal, G., Santos, A., Ledesma-Carbayo, M.J., 2010. Motion estimation of tagged cardiac magnetic resonance images using variational techniques. Computerized Medical Imaging and Graphics 34, 514–522.

Castillo, R., Castillo, E., Guerra, R., Johnson, V., McPhail, T., Garg, A., Guerrero, T., 2009. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. Physics in Medicine and Biology 54, 1849–1870.

Chui, H., Rangarajan, A., 2003. A new point matching algorithm for non-rigid registration. Computer Vision and Image Understanding 89, 114–141.

Duchi, J.C., Bartlett, P.L., Wainwright, M.J., 2012. Randomized smoothing for stochastic optimization. SIAM Journal on Optimization 22, 674–701.

Fischer, B., Modersitzki, J., 2008. Ill-posed medicine-an introduction to image registration. Inverse Problems 24, 034008–1–034008–16.

Ghosh, R.P., Horowitz-Scherer, R.A., Nikitina, T., Shlyakhtenko, L.S., Woodcock, C.L., 2010. MeCP2 binds cooperatively to its substrate and competes with histone H1 for chromatin binding sites. Molecular and cellular biology 30, 4656–4670.

Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N., 2008. Dense image registration through MRFs and efficient linear programming. Medical Image Analysis 12, 731–741.

Heckemann, R.A., Hajnal, J.V., Aljabar, P., Rueckert, D., Hammers, A., 2006. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. NeuroImage 33, 115–126.

Hollander, M., Wolfe, D.A., 1999. Nonparametric statistical methods. Wiley-Interscience.

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. NeuroImage 17, 825–841.

Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. Medical Image Analysis 5, 143–156.

Jian, B., Vemuri, B.C., 2011. Robust point set registration using gaussian mixture models. IEEE Transactions on Pattern Analysis and Machine Intelligence 33, 1633–1645.

Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P., et al., 2009a. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. NeuroImage 46, 786–802.

Klein, S., Pluim, J.P., Staring, M., Viergever, M.A., 2009b. Adaptive stochastic gradient descent optimisation for image registration. International Journal of Computer Vision 81, 227–239.

Klein, S., Staring, M., Murphy, K., Viergever, M., Pluim, J.P., et al., 2010. Elastix: a toolbox for intensity-based medical image registration. IEEE Transactions on Medical Imaging 29, 196–205.

Klein, S., Staring, M., Pluim, J.P., 2007. Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines. IEEE Transactions on Image Processing 16, 2879–2890.

Lester, H., Arridge, S.R., 1999. A survey of hierarchical non-linear medical image registration. Pattern recognition 32, 129–149.

Likar, B., Pernuš, F., 2001. A hierarchical approach to elastic registration based on mutual information. Image and Vision Computing 19, 33–44.

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P., 1997. Multimodality image registration by maximization of mutual information. IEEE Transactions on Medical Imaging 16, 187–198.

Maes, F., Vandermeulen, D., Suetens, P., 1999. Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information. Medical Image Analysis 3, 373–386.

Mandal, P.K., Mahajan, R., Dinov, I.D., 2012. Structural brain atlases: design, rationale, and applications in normal and pathological cohorts. Journal of Alzheimer's Disease 31, 169–188.

Metz, C.T., Klein, S., Schaap, M., Van Walsum, T., Niessen, W.J.,

2011. Nonrigid registration of dynamic medical imaging data using *n*D+t B-splines and a groupwise optimization approach. Medical Image Analysis 15, 238–249.

Murphy, K., Van Ginneken, B., Reinhardt, J.M., Kabus, S., Ding, K., Deng, X., Cao, K., Du, K., Christensen, G.E., Garcia, V., et al., 2011. Evaluation of registration methods on thoracic CT: the EMPIRE10 challenge. IEEE Transactions on Medical Imaging 30, 1901–1920.

Nemirovsky, A.S., Yudin, D.B., 1983. Problem complexity and method efficiency in optimization. Wiley.

Nesterov, Y., 2005. Smooth minimization of non-smooth functions. Mathematical programming 103, 127–152.

Papież, B.W., Heinrich, M.P., Fehrenbach, J., Risser, L., Schnabel, J.A., 2014. An implicit sliding-motion preserving regularisation via bilateral filtering for deformable image registration. Medical Image Analysis 18, 1299–1311.

Pluim, J.P., Maintz, J.A., Viergever, M.A., 2000. Interpolation artefacts in mutual information-based image registration. Computer Vision and Image Understanding 77, 211–232.

Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J., 1999. Nonrigid registration using free-form deformations: application to breast MR images. IEEE Transactions on Medical Imaging 18, 712–721.

Smal, I., Carranza-Herrezuelo, N., Klein, S., Wielopolski, P., Moelker, A., Springeling, T., Bernsen, M., Niessen, W., Meijering, E., 2012. Reversible jump MCMC methods for fully automatic motion analysis in tagged MRI. Medical Image Analysis 16, 301–324.

Sun, W., Niessen, W., van Stralen, M., Klein, S., 2013. Simultaneous multiresolution strategies for nonrigid image registration. IEEE Transactions on Image Processing 22, 4905–4917.

Sun, W., Niessen, W.J., Klein, S., 2014a. Free-form deformation using lower-order B-spline for nonrigid image registration, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014. Springer, pp. 194–201.

Sun, W., Niessen, W.J., Klein, S., 2014b. Randomly perturbed free-form deformation for nonrigid image registration, in: Biomedical Image Registration. Springer, pp. 62–71.

Thévenaz, P., Bierlaire, M., Unser, M., 2006. Halton sampling for image registration based on mutual information. Sampling Theory Signal Image Process 7, 141–171.

Tsao, J., 2003. Interpolation artifacts in multimodality image registration based on maximization of mutual information. IEEE Transactions on Medical Imaging 22, 854–864.

Viola, P., Wells III, W.M., 1997. Alignment by maximization of mutual information. International Journal of Computer Vision 24, 137–154.