Image registration via stochastic gradient Markov chain Monte Carlo

Daniel Grzech, Bernhard Kainz, Ben Glocker, and Loïc le Folgoc

Department of Computing, Imperial College London
{d.grzech17, b.kainz, b.glocker, l.le-folgoc}@imperial.ac.uk

Abstract. We develop a fully Bayesian framework for non-rigid registration of three-dimensional medical images, with a focus on uncertainty quantification. Probabilistic registration of large images along with calibrated uncertainty estimates is difficult for both computational and modelling reasons. To address the computational issues, we explore connections between the *Markov chain Monte Carlo by backprop* and the *variational inference by backprop* frameworks in order to efficiently draw thousands of samples from the posterior distribution. Regarding the modelling issues, we carefully design a Bayesian model for registration to overcome the existing barriers when using a dense, high-dimensional, and diffeomorphic parameterisation of the transformation. This results in improved calibration of uncertainty estimates.

1 Introduction

Image registration is the problem of aligning images into a common coordinate system such that the discrete pixel locations carry the same semantic information. It is a common pre-processing step for many applications, *e.g.* the statistical analysis of imaging data and computer-aided diagnosis. Image registration methods based on deep learning tend to incorporate task-specific knowledge from large datasets [3], whereas traditional methods are more general purpose [11]. Many established models [9,14,11] are based on the iterative optimisation of an energy function consisting of task-specific similarity and regularisation terms, which leads to an estimated deformation field and has to be done independently for every pair of images to be registered.

VoxelMorph [3,2,6,7] changed this paradigm by learning a function that maps a pair of input images to a deformation field. This gave a speed-up of several orders of magnitude while maintaining an accuracy comparable to established methods. An overview of current learning-based methods for registration can be found in [16]. With a few notable exceptions [6,7], Bayesian methods are often shunned when designing novel medical image analysis algorithms because of their perceived conceptual challenges and computational overhead. Yet in order to fully explore the parameter space and to lessen the impact of ad-hoc hyperparameter choices, it is desirable to adopt a Bayesian point of view.

Markov chain Monte Carlo (MCMC) methods have been used for asymptotically exact sampling from the posterior distribution in rigid registration [13],

and are popular for analysing non-rigid registration uncertainty in intra-subject studies [20]. Recent research shows that the computational burden of MCMC can be lessened by embedding it in a multilevel framework [21]. The problem of uncertainty quantification has also been addressed using variational Bayesian methods [22]. In [15] the authors compared the quality of uncertainty estimates from an efficient and approximate variational Bayesian model and a reversible jump MCMC model, which is asymptotically exact.

We use the stochastic gradient Markov chain Monte Carlo (SG-MCMC) algorithm to establish an efficient posterior sampling algorithm for non-rigid image registration. SG-MCMC is based on the idea of stochastic gradient descent interpreted as a stochastic process with a stationary distribution centred on the optimum and with a covariance structure that can be used to approximate the posterior distribution [5,18]. The following is the summary of our main contributions:

1. We propose an efficient SG-MCMC algorithm for three-dimensional diffeomorphic non-rigid image registration;
2. We propose a new regularisation loss, which allows to carry out inference of the regularisation strength in a setting with a very high number of degrees of freedom (d.f.);
3. We evaluate the performance of our model both qualitatively and quantitatively by analysing the output uncertainty estimates on inter-subject brain MRI data.

To our knowledge, this is the first time that SG-MCMC has been used for the task of image registration. The code is available in a public repository: https://github.com/dgrzech/ir-sgmcmc.

Related work. Bayesian parameter estimation for established registration models was proposed in [27]. Bayesian frameworks have been used to characterize image intensities [10] and anatomic variability [26]. Kernel regression has also been used to tackle multi-modal image registration with uncertainty [12,28]. We believe that our work is the first that efficiently tackles Bayesian image registration and uncertainty estimation using a very high-dimensional parameterisation of the transformation.

2   Registration model

We denote an image pair by $\mathcal{D} = (F, M)$, where $F : \Omega_F \to \mathbb{R}$ is a fixed image and $M : \Omega_M \to \mathbb{R}$ is a moving image. We assume that F can be generated from M if deformed by a transformation $\varphi : \Omega_F \to \Omega_M$ which is parameterised by w. The goal of registration is to align the underlying domains Ω_F and Ω_M using a mapping that roughly visually aligns the images F and $M(w) \coloneqq M \circ \varphi^{-1}(w)$ and is physically plausible, *i.e.* find parameters w such that $F \simeq M(w)$. We parameterise the transformation using the stationary velocity field (SVF) formulation. The velocity field is integrated numerically through scaling-and-squaring which results in a diffeomorphic transformation [1].

Likelihood model. The likelihood model $p\left(\mathcal{D} \mid w\right)$ specifies the relationship between the data and the transformation parameters through the choice of a similarity metric. Due to its robustness to linear intensity transformations we use a similarity metric based on local cross-correlation (LCC). However, because LCC is not meaningful in a probabilistic context, we opt for the sum of voxel-wise squared differences instead of the usual sum of the voxel-wise product of intensities. Thus we can also enhance the likelihood model with extra features.

Denote the fixed and the warped moving images, with intensities standardised to zero mean and unit variance inside a neighbourhood of 3 voxels, as $\overline{F}$ and $\overline{M(w)}$ respectively. Following the example in [15], in order to make the model more robust to high outlier values caused by acquisition artifacts and misalignment over the course of registration, we adopt a Gaussian mixture model (GMM) of intensity residuals. At voxel k, the corresponding intensity residual r_k is assigned to the l-th component of the mixture, $1 \leqslant l \leqslant L$, if the categorical variable $c_k \in \{1, \cdots, L\}$ is equal to l. It then follows a normal distribution $\mathcal{N}(0, \beta_l^{-1})$. The component assignment c_k follows a categorical distribution and takes value l with probability ρ_l. In all experiments we use $L = 4$ components.

We also use virtual decimation to account for the fact that voxel-wise residuals are not independent, preventing over-emphasis on the data term and allowing to better calibrate uncertainty estimates [23]. The full expression of the image similarity term is then given by:

$$\mathcal{E}_{\text{data}} = \alpha \times - \sum_{i=1}^{N} \log \sum_{l=1}^{L} \frac{\sqrt{\beta_l}}{\sqrt{2\pi}} \rho_l \exp\left(-\frac{\beta_l}{2}\|\overline{F} - \overline{M(w)}\|^2\right) \qquad (1)$$

where α is the scalar virtual decimation factor.

Transformation priors. In Bayesian models, the transformation parameters are typically regularised with use of a multivariate normal prior $p(w \mid \lambda) = |\lambda L^T L|^{\frac{1}{2}} (2\pi)^{-\frac{N}{2}} \exp^{-\frac{1}{2}\lambda(Lw)^T Lw}$ that ensures smoothness, where N is the number of voxels in the image, λ is a scalar parameter that controls the strength of regularisation, and L is the matrix of a differential operator, here chosen to penalise the magnitude of the first derivative of the velocity field. Note that $(Lw)^T Lw = \|Lw\|^2$.

The regularisation strength parameter λ can be either fixed [3] or learnt from the data. The latter has been done successfully only in the context of transformation parameterisations with a relatively low number of d.f., *e.g.* B-splines [23] or a sparse learnable parameterisation [15]. In case of an SVF, where the number of d.f. is orders of magnitude higher, the problem is even more difficult. The baseline method that we use for comparison with our proposed regularisation loss, which was described in [23], corresponds to an uninformative gamma prior.

In order to infer the regularisation strength we specify a prior on the scalar regularisation energy $\chi^2 := \|Lw\|^2$. We use a log-normal prior on χ^2 and derive

a prior on the velocity field:

$$\log p(\chi^2) \propto \log \chi^2 + \log \sigma_{\chi^2} + \frac{(\log \chi^2 - \mu_{\chi^2})^2}{2\sigma_{\chi^2}^2} \tag{2}$$

$$\log p(w) \propto \log p(\chi^2) + (\frac{\nu}{2} - 1) \cdot \log \chi^2 \tag{3}$$

where $\nu = N \cdot 3$ is the number of d.f. Given semi-informative hyperpriors on μ_{χ^2} and $\sigma_{\chi^2}^2$, which we discuss in the next section, we can estimate the right regularisation strength from data. Overall, the regularisation term is given by $\mathcal{E}_{\mathrm{reg}} = \log p(\chi^2) + \log p(w)$.

Hyperpriors. Parameters of the priors are treated as latent variables. We set the likelihood model hyperpriors similarly to [15], with the parameters β_l assigned independent log-normal priors $\mathrm{Lognormal}(\beta_l \mid \mu_{\beta_0}, \sigma_{\beta_0}^2)$ and the mixture proportions $\rho = (\rho_1, \cdots, \rho_L)$ with an uninformative Dirichlet prior $\mathrm{Dir}(\rho \mid \kappa)$, where $\kappa = (\kappa_1, \cdots, \kappa_L)$. The problem of inferring regularisation strength is difficult, so we use semi-informative priors for the transformation prior parameters. The exponential of the transformation prior parameter μ_{χ^2} follows a gamma distribution $\Gamma(\exp\left(\mu_{\chi^2}\right) \mid a_{\chi_0^2}, b_{\chi_0^2})$ and $\sigma_{\chi^2}^2$ has a log-normal prior $\mathrm{Lognormal}(\sigma_{\chi^2}^2 \mid \mu_{\chi_0^2}, \sigma_{\chi_0^2}^2)$.

3   Variational inference

To initialise the MCMC algorithm we use the result of variational inference (VI). We assume that the approximate posterior distribution of the transformation parameters $q_w \sim \mathcal{N}(\mu_w, \Sigma_w)$ is a multivariate normal distribution. Due to the dimensionality of the problem, computing the full covariance matrix is not possible, so we model it as a sum of diagonal and low-rank parts $\Sigma_w = \mathrm{diag}(\sigma_w^2) + u_w u_w^T$, with σ_w^2 and u_w both of size $N \cdot 3 \times 1$. To carry out VI, we maximise the evidence lower bound (ELBO), which can be written as:

$$\mathcal{L}(q) = \mathbb{E}_q\left[\log p(\mathcal{D} \mid w)\right] - \mathrm{D_{KL}}(q \mid\mid p) = -\langle\mathcal{E}_{\mathrm{data}} + \mathcal{E}_{\mathrm{reg}}\rangle_q + H(q) \tag{4}$$

where $\mathrm{D_{KL}}(q \mid\mid p)$ is the Kullback-Leibler divergence between the approximate posterior q and the prior p. This corresponds to the sum of similarity and regularisation terms, with an additional term equal to the entropy of the posterior distribution $H(q)$. We use the reparameterisation trick with two samples per update to backpropagate w.r.t. parameters of the approximate variational posterior q_w, i.e. μ_w, σ_w^2, and u_w.

In order to make optimisation less susceptible to undesired local minima we take advantage of Sobolev gradients [19]. Samples from q_w are convolved with a Sobolev kernel. To lower the computational cost, we approximate the 3D kernel by three separable 1D kernels [24].

4   Stochastic gradient Markov chain Monte Carlo

We use stochastic gradient Langevin dynamics (SGLD) [4,25] to sample the transformation parameters in an efficient way:

$$w_{k+1} \leftarrow w_k + \tau \sigma_w^2 \nabla \log q(w_k) + \sqrt{2\tau} \sigma_w \xi_k \qquad (5)$$

where τ is the step size, $\nabla \log q(w_k)$ is an estimate of the gradient of the posterior probability density function, and ξ_k is an independent draw from a multivariate normal distribution with zero mean and an identity covariance matrix.

Given a sufficient number of steps SGLD puts no restrictions on how the chain is initialised, but in order to lower the mixing time we set $w_0 \leftarrow \mu_w$. In the limit as $\tau \to 0$ and $k \to \infty$, it allows for asymptotically exact sampling from the posterior of the transformation parameters. The scheme suffers from similar issues as Gibbs sampling used in [15], *i.e.* high autocorrelation and slow mixing between modes. On the other hand, the term corresponding to the gradient of the posterior probability density function allows for more efficient landscape transversal. Moreover, simplicity of the formulation makes SGLD better suited to a high-dimensional problem like image registration.

The value of τ is important here and should be smaller than the width of the most constrained direction in the local energy landscape, which can be estimated using Σ_w. We discard the first 2,000 samples output by the algorithm to allow for the chain to reach the stationary distribution.

5   Experiments

The model was implemented in PyTorch. For all experiments we use three-dimensional brain MRI scans from the UK Biobank dataset. Input images were resampled to 96^3 voxels, with isotropic voxels of length 2.43 mm, and registered with the affine component of *drop2* [8]. Note that the model is not constrained by memory, so it can be run on higher resolution images to produce output that is more clinically relevant, while maintaining a high speed of sampling.

We use the Adam optimiser with a learning rate of 5×10^{-3} for VI and the SGD optimiser with a learning rate of 1×10^{-1} for SG-MCMC. The hyperprior parameters are set to $\mu_{\beta_o} = 0$, $\sigma_{\beta_0}^2 = 2.3$, $\kappa = 0.5$, $a_{\chi_0^2} = 0.5 \cdot \nu$, $b_{\chi_0^2} = 0.5 \cdot \lambda_0$, $\mu_{\chi_0^2} = 2.8$, and $\sigma_{\chi_0^2}^2 = 5$, where λ_0 is the desired strength of equivalent $L2$ regularisation at initialisation. The model is particularly sensitive to the value of the transformation prior parameters. We start with an identity transformation, σ_w of half a voxel in each direction, and u_w set to zero, and VI is run until the loss value plateaus. We are unable to achieve convergence in the sense of the magnitude of updates to Σ_w.

Regularisation strength. In the first experiment we show the benefits of the proposed regularisation loss. We compare the output of VI when using a fixed regularisation weight $\lambda \in \{0.01, 0.1\}$, the baseline method for learnable regularisation strength, and the novel regularisation loss. The result is shown

in Figure 1. The output transformation is highly sensitive to the regularisation weight and so is registration uncertainty, hence the need for a reliable method to infer regularisation strength from data.

Fixed image F and displacement corresp. μ_w

Warped moving image $M(\mu_w)$

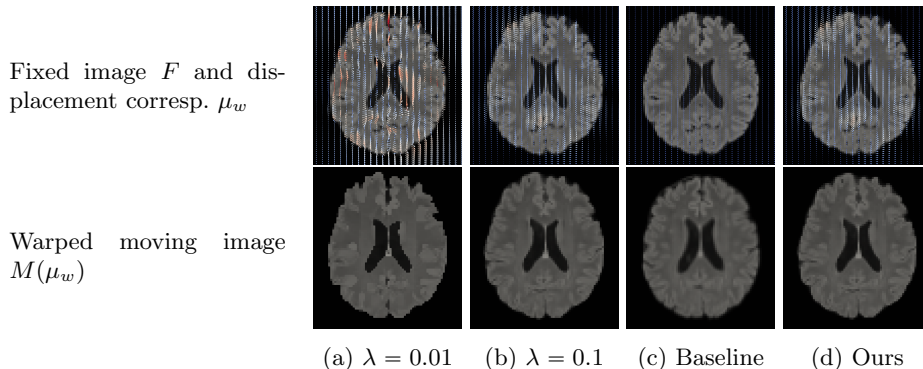(a) $\lambda = 0.01$ (b) $\lambda = 0.1$ (c) Baseline   (d) Ours

Fig. 1: Output when using a fixed regularisation weight, the baseline method for learnable regularisation strength, and our regularisation loss. For the baseline, the regularisation strength is so high that it prevents alignment of the images, showing that the existing schemes for inferring regularisation strength from data are inadequate in cases with a very large number of d.f. Middle slice of 3D images in the axial plane.

In Figure 2 we show the output of VI for two pairs of images which require different regularisation strengths. We choose a fixed image F, two moving images M_1 and M_2, and two regularisation weights $\lambda \in \{0.1, 0.4\}$. Use of our regularisation loss, which at initialisation corresponds to $\lambda = 0.4$, prevents over-smoothing. Due to its characteristics, it is preferable to initialise its strength to a higher value.

Uncertainty quantification. To evaluate registration uncertainty we calculate the mean and the standard deviation of displacement using 50 samples selected at random from the output of SG-MCMC. Figure 3 shows the result for a pair of input images. In order to assess the results quantitatively, we use subcortical structure segmentations. We calculate Dice scores (DSC) and mean surface distances (MSD) between the fixed segmentation and the moving segmentation warped with the mean transformation, and compare them to those obtained using the 50 sample transformations. We report these metrics in Table 1 and Figure 3.

The statistics prove that the posterior samples output by SG-MCMC are of high quality and varied. For a small number of structures (the left and right accumbens and thalamus, and the right caudate) the metrics are minimally worse for the mean transformation than before non-rigid registration. In case of the thalamus this can be attributed to a sub-optimal regularisation strength. The registration error for the accumbens and the caudate is likely caused by their

| | (a) $\lambda = 0.1$ | (b) $\lambda = 0.4$ | (c) Ours |
| M_1 | | | |

Fixed image F and displacement corresp. μ_w

Warped moving image $M_1(\mu_w)$

| | (d) $\lambda = 0.1$ | (e) $\lambda = 0.4$ | (f) Ours |
| M_2 | | | |

Fixed image F and displacement corresp. μ_w
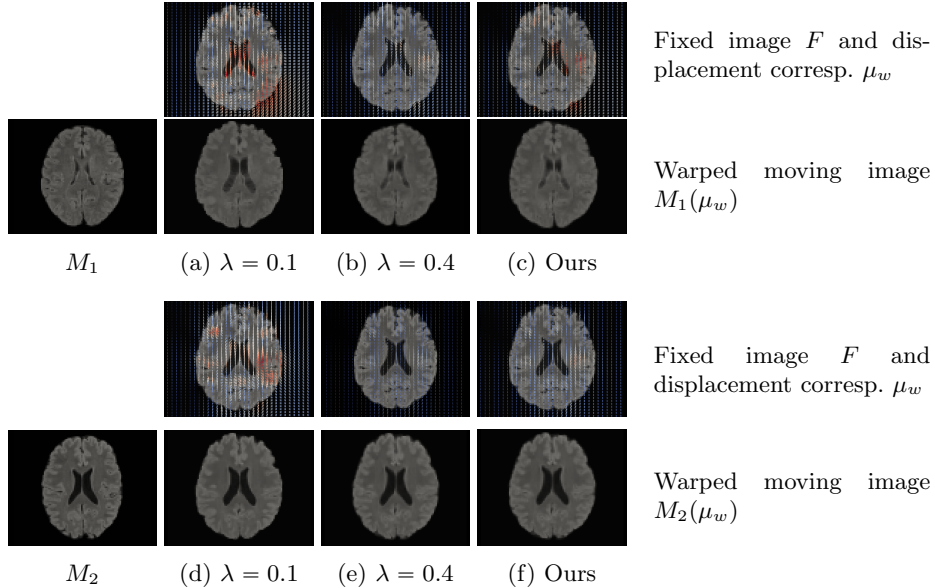
Warped moving image $M_2(\mu_w)$

Fig. 2: Output of VI for two pairs of images which require different regularisation strengths. At initialisation the strength of our loss corresponds to the fixed regularisation weight $\lambda = 0.4$. Middle slice of 3D images in the axial plane.

Table 1: DSC and MSD for a number of subcortical structures pre-registration and after applying the mean transformation calculated from the output of SG-MCMC.

structure	DSC			MSD (mm)		
	before	mean	SD	before	mean	SD
brain stem	0.815	0.879	0.002	1.85	1.17	0.03
L/R accumbens	0.593/0.653	0.637/0.592	0.036/0.022	1.20/1.13	1.03/1.18	0.13/0.10
L/R amygdala	0.335/0.644	0.700/0.700	0.019/0.015	2.18/1.44	1.12/1.12	0.08/0.08
L/R caudate	0.705/0.813	0.743/0.790	0.011/0.008	1.37/1.44	1.21/0.99	0.05/0.06
L/R hippocampus	0.708/0.665	0.783/0.781	0.009/0.009	1.45/1.60	1.00/1.03	0.05/0.05
L/R pallidum	0.673/0.794	0.702/0.798	0.014/0.014	1.56/1.12	1.29/0.98	0.07/0.08
L/R putamen	0.772/0.812	0.835/0.856	0.007/0.006	1.30/1.02	0.92/0.78	0.05/0.05
L/R thalamus	0.896/0.920	0.881/0.901	0.005/0.004	0.90/0.67	0.92/0.86	0.04/0.05

tiny size. Thus the label distribution appears credible in the sense defined in [17]. The output is also consistent with previous findings on registration uncertainty, *e.g.* higher uncertainty in homogenous regions [23].

(a) Fixed image   (b) Moving image   (c) Mean   (d) SD magnitude
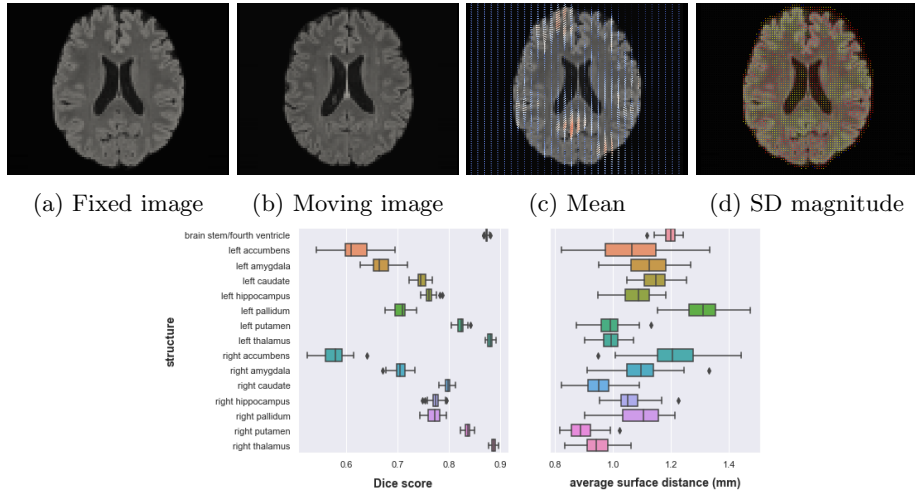

Fig. 3: Output of SG-MCMC for a pair of input images, calculated using 50 samples.

6   Discussion

Modelling assumptions. The quality of uncertainty estimates is sensitive to the initialisation of regularisation loss hyperparameters and the validity of model assumptions. These include: 1. coinciding image intensities up to the expected spatial noise offsets, 2. ignoring spatial correlations between residuals, and 3. the spherical covariance structure of the approximate posterior in VI. The first assumption is valid in case of mono-modal registration but the model can be easily adapted to other settings by changing the data loss. In future work we plan to the use a frequency-domain model to deal with the last assumption.

Implementation and computational efficiency. The experiments were run on a system with an Intel i9-7900X CPU and a GeForce RTX 2080Ti GPU. VI took approx. 5 min per image pair and SG-MCMC produced 5 samples per second. Due to lack of data it is difficult to directly compare the runtime with that of other Bayesian image registration methods, but it is an order of magnitude better than in other recent work [15], while also being three- rather than two-dimensional.

7   Conclusion

In this paper we present an efficient Bayesian model for three-dimensional medical image registration. The newly proposed regularisation loss allows to tune the regularisation strength using a parameterisation of transformation that involves a very large number of d.f. Sampling from the posterior distribution via SG-MCMC makes it possible to quantify registration uncertainty for high-resolution images.

References

1. Arsigny, V., Commowick, O., Pennec, X., Ayache, N.: A Log-Euclidean Framework for Statistics on Diffeomorphisms. In: MICCAI. pp. 924–931 (2006)
2. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: An Unsupervised Learning Model for Deformable Medical Image Registration. In: CVPR (2018)
3. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: VoxelMorph: A Learning Framework for Deformable Medical Image Registration. IEEE Transactions on Medical Imaging (2019)
4. Besag, J.: Comments on "Representations of knowledge in complex systems" by U. Grenander and MI Miller. Journal of the Royal Statistical Society **56**, 591–592 (1993)
5. Chen, C., Carlson, D., Gan, Z., Li, C., Carin, L.: Bridging the Gap between Stochastic Gradient MCMC and Stochastic Optimization. In: AISTATS (2016)
6. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R.: Unsupervised Learning for Fast Probabilistic Diffeomorphic Registration. In: MICCAI (2018)
7. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R.: Unsupervised Learning of Probabilistic Diffeomorphic Registration for Images and Surfaces. MedIA: Medical Image Analysis **57**, 226–236 (2019)
8. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registration through mrfs and efficient linear programming. Medical image analysis **12**(6), 731–741 (2008)
9. Glocker, B., Sotiras, A., Komodakis, N., Paragios, N.: Deformable medical image registration: setting the state of the art with discrete methods. Annual review of biomedical engineering **13**, 219–244 (2011)
10. Hachama, M., Desolneux, A., Richard, F.J.: Bayesian technique for image classifying registration. IEEE Transactions on Image Processing **21**(9), 4080–4091 (2012)
11. Heinrich, H.P., Jenkinson, M., Brady, M., Schnabel, J.A.: MRF-Based Deformable Registration and Ventilation Estimation of Lung CT. IEEE Transactions on Medical Imaging **32**(7), 1239–1248 (7 2013)
12. Janoos, F., Risholm, P., Wells, W.: Bayesian characterization of uncertainty in multi-modal image registration. In: Dawant, B.M., Christensen, G.E., Fitzpatrick, J.M., Rueckert, D. (eds.) Biomedical Image Registration. pp. 50–59. Springer Berlin Heidelberg, Berlin, Heidelberg (2012)
13. Karabulut, N., Erdil, E., Çetin Müjdat: A Markov Chain Monte Carlo based Rigid Image Registration Method. Tech. rep. (2017)
14. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.: Elastix: a toolbox for intensity-based medical image registration. IEEE Transactions on Medical Imaging **29**(1), 196–205 (2009)
15. Le Folgoc, L., Delingette, H., Criminisi, A., Ayache, N.: Quantifying Registration Uncertainty With Sparse Bayesian Modelling. IEEE Transactions on Medical Imaging **36**(2), 607–617 (2017)
16. Lee, M.C.H., Oktay, O., Schuh, A., Schaap, M., Glocker, B.: Image-and-Spatial Transformer Networks for Structure-Guided Image Registration. In: MICCAI (2019)

17. Luo, J., Sedghi, A., Popuri, K., Cobzas, D., Zhang, M., Preiswerk, F., Toews, M., Golby, A., Sugiyama, M., Wells Iii, W.M., Frisken, S.: On the Applicability of Registration Uncertainty. In: MICCAI (2019)
18. Mandt, S., Hoffman, M.D., Blei, D.M.: Stochastic Gradient Descent as Approximate Bayesian Inference. Journal of Machine Learning Research **18**, 1–35 (2017)
19. Neuberger, J.W., Dold, A., Takens, F.: Sobolev Gradients and Differential Equations. Lecture Notes in Computer Science, Springer (1997)
20. Risholm, P., Janoos, F., Norton, I., Golby, A.J., Wells, W.M.: Bayesian characterization of uncertainty in intra-subject non-rigid registration. Medical Image Analysis (2013)
21. Schultz, S., Handels, H., Ehrhardt, J.: A multilevel Markov Chain Monte Carlo approach for uncertainty quantification in deformable registration. In: SPIE Medical Imaging (2018)
22. Schultz, S., Krüger, J., Handels, H., Ehrhardt, J.: Bayesian inference for uncertainty quantification in point-based deformable image registration. In: SPIE Medical Imaging. p. 46. No. March 2019 (2019)
23. Simpson, I.J., Schnabel, J.A., Groves, A.R., Andersson, J.L., Woolrich, M.W.: Probabilistic inference of regularisation in non-rigid registration. Neuroimage **59**(3), 2438–2451 (Feb 2012)
24. Slavcheva, M., Baust, M., Ilic, S.: SobolevFusion: 3D Reconstruction of Scenes Undergoing Free Non-rigid Motion. In: CVPR (2018)
25. Welling, M., Teh, Y.W.: Bayesian Learning via Stochastic Gradient Langevin Dynamics. In: ICML. pp. 681–688 (2011)
26. Zhang, M., Fletcher, P.T.: Bayesian principal geodesic analysis in diffeomorphic image registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 121–128. Springer (2014)
27. Zhang, M., Singh, N., Fletcher, P.T.: Bayesian estimation of regularization and atlas building in diffeomorphic image registration. In: International conference on information processing in medical imaging. pp. 37–48. Springer (2013)
28. Zöllei, L., Jenkinson, M., Timoner, S., Wells, W.: A marginalized map approach and em optimization for pair-wise registration. In: Biennial International Conference on Information Processing in Medical Imaging. pp. 662–674. Springer (2007)