A Bayesian Approach for Spatially Adaptive Regularisation in Non-rigid Registration

Ivor J.A. Simpson[1,2,⋆], Mark W. Woolrich[3], Manuel Jorge Cardoso[1], David M. Cash[1], Marc Modat[1], Julia A. Schnabel[2], and Sebastien Ourselin[1]

[1] Centre for Medical Image Computing, University College London
[2] Institute of Biomedical Engineering, University of Oxford
[3] Oxford Centre for Human Brain Activity, University of Oxford

Abstract. This paper introduces a novel method for inferring spatially varying regularisation in non-rigid registration. This is achieved through full Bayesian inference on a probabilistic registration model, where the prior on transformations is parametrised as a weighted mixture of spatially localised components. Such an approach has the advantage of allowing the registration to be more flexibly driven by the data than a more traditional global regularisation scheme, such as bending energy. The proposed method adaptively determines the influence of the prior in a local region. The importance of the prior may be reduced in areas where the data better supports deformations, or can enforce a stronger constraint in less informative areas. Consequently, the use of such a spatially adaptive prior may reduce the unwanted impact of regularisation on the inferred deformation field. This is especially important for applications such as tensor based morphometry, where the features of interest are directly derived from the deformation field. The proposed approach is demonstrated with application to tensor based morphometry analysis of subjects with Alzheimer's disease and healthy controls. The results show that using the proposed spatially adaptive prior leads to deformation fields that have a substantially lower average complexity, but which also provide more accurate localisation of statistical group differences.

1 Introduction

Non-rigid image registration is an important tool for analysing morphometric differences in subjects with pathology, such as Alzheimer's disease (AD), from healthy controls using structural magnetic resonance (MR) images of the brain. However, non-rigid registration is an ill-posed problem. The ambiguity in intensity matching makes it infeasible for the data alone to provide a unique voxelwise mapping. Instead, a reasonable mapping between images is sought after, which maximises the similarity of appearance whilst maintaining a plausible deformation. This is achieved through the use of regularisation, which constrains the magnitude and smoothness of the deformation.

⋆ I. Simpson and M. Modat were supported by the CBRC Strategic Investment Award (Ref. 168) and the the NIHR Queen Square Dementia BRU. S. Ourselin and M.J. Cardoso were supported by EPSRC (EP/H046410/1).

K. Mori et al. (Eds.): MICCAI 2013, Part II, LNCS 8150, pp. 10–18, 2013.

Tensor based morphometry (TBM) uses non-rigid registration to describe morphological differences [1]. TBM involves analysing the Jacobian tensor of deformation fields, which are derived through the registration of a population of images to an atlas. TBM offers a whole brain approach to statistical analysis, and has the potential to extract rich features that accurately summarise anatomical differences. A limitation of TBM is that the inferred morphological features suffer from a dependence on the registration algorithm, and its parametrisation. In particular, the choice of the form and scale of regularisation has a strong impact on the deformation field. This paper proposes a novel spatially adaptive regularisation model and inference scheme that helps to alleviate this issue.

Regularisation can be thought of as a prior on the set of expected deformations. Commonly, these priors have a physical interpretation e.g. bending energy, fluid or linear elasticity. However, these models are most commonly described as having the same level of effect across the image. This assumption may be unreasonable, as different regions of the image contain different amount of information. Uninformative, or noisy image areas should be strongly influenced by the priors, whereas feature rich regions should be given more freedom. Furthermore, the level of anatomical mis-correspondence is also likely to vary across space, and some regions may require more complex deformations than others.

Previous work on the use of spatially varying regularisation includes approaches that vary across different segmented regions [2][3][4]. More data driven approaches include the anisotropic smoothing of image similarity gradients, according to image information [5]. These approaches are somewhat ad-hoc, as the overall scale of the regularisation and data fidelity needs to be defined a-priori. Risholm et al. [6] present a Bayesian inference scheme that allows the linear elastic parameters to be inferred spatially. However, the vast computational complexity of the inference scheme limits its general application.

This paper proposes a model and tractable Bayesian inference strategy that allows spatially varying regularisation. Registration is described as a probabilistic model, with a transformation prior that is parametrised by a set of hyper-parameters. Each hyper-parameter influences a spatially localised region of the prior. Through the use of full Bayesian inference, the hyper-parameter weights can be inferred alongside the transformation. This allows the effects of the prior to be spatially adaptively determined depending on the images being registered.

The novelty of this work lies in the inference of spatially varying regularisation for non-rigid registration. This model is demonstrated for TBM using subjects with AD and healthy controls. The proposed method out-performs either an adaptive, or fixed strength global regularisation model in terms of feature localisation, and provides deformation fields with more reasonable complexity.

2 Method

2.1 Model

Image registration can be described probabilistically using a generative model, where it is assumed that the target image, $\mathbf{y}$, can be generated from a source image, $\mathbf{x}$, when it is deformed according to some transformation model. Here, $\mathbf{t}(\mathbf{x}, \mathbf{w})$ is the transformed source image, where $\mathbf{w}$ parametrises the transformation. In this work the cubic B-spline is used as the transformation model and $\mathbf{w}$ refers to the displacements of the B-spline control points.

This model includes a noise term, $\mathbf{e}$, that describes the accuracy of model fit. Here, $\mathbf{e}$ is assumed to be independently and identically distributed across voxels, and follows a Normal distribution, $\mathbf{e} \approx \mathcal{N}(0, \mathbf{I}\phi^{-1}\alpha)$. $\mathbf{I}$ is an identity matrix the size of the number of voxels, N_v and ϕ corresponds to noise precision. The intensities of the residual image are not truly independent, as they feature a level of spatial covariance. This is compensated for by α, which is a virtual decimation factor [7]. The full generative model is therefore given as: $\mathbf{y} = \mathbf{t}(\mathbf{x}, \mathbf{w}) + \mathbf{e}$.

2.2 Priors

Prior information can be used to constrain the parameters of the model. The noise in model fit, ϕ, is well defined by the data, so an uninformative Gamma prior is used, $\phi = \text{Ga}(a_0, b_0)$. However, the transformation parameters, $\mathbf{w}$, require a more informative constraint, and ought to encourage smooth and small deformations, unless adequately justified by the data.

A prior on $\mathbf{w}$ can be described using a multivariate Normal distribution, $P(\mathbf{w}) = \mathcal{N}(0, \mathbf{\Sigma})$. Here, the mean of the prior is 0, which represents the identity transformation. $\mathbf{\Sigma}$ describes the expected variance, and covariance of the transformation parameters. This definition allows the specification of complex and flexible priors. Commonly, bending, or linear elastic energy priors have been encoded in such a form [8]. Conversely, weakly informative constraints could also be encoded, such as the magnitude of the deformation parameters (L1 norm).

In this work, we adopt an approach similar to the multiple sparse priors (MSP) of Friston et al. [9], which was demonstrated for the M/EEG inverse problem. The MSP method was chosen as it allows a flexible definition of prior components. In this approach, the prior covariance matrix is defined as a weighted mixture of covariance components: $\mathbf{\Sigma} = \sum_i^n \lambda_i \Sigma_i$, where λ_i corresponds to the weight and Σ_i is a spatially localised prior component.

The prior components are derived from sparsely sampled columns of a spatial coherence prior, G. Here, G is a squared exponential Gaussian process prior, and can equivalently be considered a Green's function of a discrete diffusion process defined on a transformation parameter adjacency matrix, A. G is defined as: $G(\sigma) = \exp(\sigma A) \approx \sum_i^4 \frac{\sigma^i}{i!} A^i$, where $A_{i,j}$ is 1 where the parameters are adjacent, and 0 everywhere else. The parameter σ controls the local coherence between nearby transformation parameters. This approximation to the Green's function only accounts for 4th order neighbouring transformation parameters, allowing priors with compact spatial support.

A given prior component, i, is defined as: $\Sigma_i = q_i q_i^T$, where q_i corresponds to the sampled column from $G(\sigma)$. In the present model, each prior component, Σ_i, controls the variance of one control point displacement in a direction, and the covariance with nearby control points. The scale of this variance/covariance is dictated by its control parameter λ_i, which is inferred from the data. In the present model, there is a uniform prior on λ_i and $\{\lambda\} = \{\lambda_1, \lambda_2, ..., \lambda_{N_c}\}$, where N_c is the number of transformation parameters.

2.3 Inference

Numerical integration approaches are computationally prohibitive in problems with many parameters. For this reason, variational Bayes [10] (VB) was chosen as the inference strategy. VB allows tractable full Bayesian inference, and has been demonstrated for use in high resolution non-rigid registration [7].

VB approximates the posterior distribution of model parameters as parametric distributions. In this work, mean-field VB provides the inference, hence the posterior distribution on the model parameters is approximated as:

$$P(\mathbf{w}, \phi, \lambda|\mathbf{y}) \approx q(\mathbf{w}, \phi, \{\lambda\}) \approx q(\mathbf{w})q(\phi)\prod_{i}^{n} q(\lambda_i) \quad (1)$$

The functional forms of the approximate posterior distributions are constrained to be conjugate to the priors, and are given as: $q(\mathbf{w}) = \mathcal{N}(\mathbf{w}; \mu, \mathbf{\Upsilon})$ and $q(\phi) = Ga(\phi, a, b)$. Where a single global λ is used, this can also be described as a Gamma distribution. In this work, only point estimates are obtained for $\{\lambda\}$.

Subject to these approximations, the Bayesian cost function is the negative variational free energy, $\mathcal{F}$. $\mathcal{F}$ is a sum of the marginal log-likelihood, with respect to the approximate posterior parameter distributions, and the Kullback-Leibler divergence between the prior and posterior distributions. Maximising $\mathcal{F}$ minimises the distance between the approximate and the true posterior distributions. Through the calculus of variations, analytic updates can be found for the parameters of the approximate posterior distributions $q(\mathbf{w})$ and $q(\phi)$:

$$\mathbf{\Upsilon} = (\alpha\bar{\phi}\mathbf{J}^\mathsf{T}\mathbf{J} + \mathbf{\Sigma}^{-1})^{-1} \quad (2) \qquad \mu_{new} = \mathbf{\Upsilon}\left[\alpha\bar{\phi}\mathbf{J}^\mathsf{T}(\mathbf{J}\mu_{old} + \mathbf{k})\right] \quad (4)$$

$$b = b_0 + \frac{N_v\alpha}{2} \quad (3) \qquad \frac{1}{a} = \frac{1}{a_0} + \frac{1}{2}\alpha(\mathbf{k}^\mathsf{T}\mathbf{k} + \mathrm{Trace}(\mathbf{\Upsilon}\mathbf{J}^\mathsf{T}\mathbf{J})) \quad (5)$$

where $\mathbf{J}$ is the matrix of first order partial derivatives of the transformation parameters with respect to the transformed image $\mathbf{t}(\mathbf{x}, \mu_{old})$. μ_{old} and μ_{new} describe the previous and new estimates of the mean transformation parameters. $\mathbf{k}$ is the vector representing the residual image $\mathbf{y} - \mathbf{t}(\mathbf{x}, \mathbf{w})$. $\mathbf{\Upsilon}$ is the estimated posterior covariance matrix of $\mathbf{w}$. $\bar{\phi}$ is the expectation of the estimated noise precision.

Inferring the Spatial Prior. Point estimates of $\{\lambda\}$ are optimised with respect to $\mathcal{F}$. The terms of $\mathcal{F}$ that relate to $\mathbf{\Sigma}$ are given below:

$$\mathcal{F} = \frac{1}{2}\left(-\log|\mathbf{\Sigma}| - \mathrm{Trace}(\mathbf{\Upsilon}\mathbf{\Sigma}^{-1}) - \mu\mathbf{\Sigma}^{-1}\mu\right) + const\{\mathbf{\Sigma}\} \quad (6)$$

As λ only directly influences $\mathbf{\Sigma}$, the derivative of $\mathcal{F}$ with respect to each local regularisation control parameter, λ_i, can therefore be expressed as:

$$\frac{\partial \mathcal{F}}{\partial \lambda_i} = \frac{1}{2}\left[-\text{Trace}\left(\mathbf{\Upsilon}\frac{\partial \mathbf{\Sigma}^{-1}}{\partial \lambda_i}\right) + \text{Trace}\left(\mathbf{\Sigma}\frac{\partial \mathbf{\Sigma}^{-1}}{\partial \lambda_i}\right) - \mu^T \frac{\partial \mathbf{\Sigma}^{-1}}{\partial \lambda_i}\mu\right] \tag{7}$$

where $\frac{\partial \mathbf{\Sigma}^{-1}}{\partial \lambda_i} = -\mathbf{\Sigma}^{-1}\lambda_i \Sigma_i \mathbf{\Sigma}^{-1}$. The Bayesian cost function optimises λ to balance three components: The first trace term prefers the prior to be similar to the posterior covariance, $\mathbf{\Upsilon}$. The second trace term reduces the variance of the prior. The final term reduces the penalisation of the distance of μ from the identitiy transformation. The step size along $\frac{\mathcal{F}}{\lambda_i}$ is chosen such that the mean change in $\mathbf{\Sigma}$ is equivalent to the change given by optimising a global λ using VB [7]. The parameters $q(\mathbf{w})$, $q(\phi)$ and $\{\lambda\}$ are alternately optimised to fit the model. $\{\lambda\}$ is initialised to give a prior transformation variance of 5mm.

3 Materials

170 structural MR images acquired on a 3T scanner were taken from the ADNI2[1] database. 85 of these subjects suffered from AD, the other 85 are healthy controls (HC). These subjects were chosen based on the 85 currently available AD subjects in ADNI2, and the first 85 HC ranked numerically based on subject ID.

4 Experiments

4.1 Groupwise Atlas Construction

All subjects were initially rigidly registered to the MNI 152 atlas, and averaged. This was used as a starting point for an iterative groupwise average [11]. The atlas creation process consisted of 5 iterations of affine registrations, and 5 iterations of non-rigid registrations. All the non-rigid registrations were run in a hierarchical scheme, and the final B-spline knot spacing was 10mm.

The groupwise image was created using a fixed level of bending energy regularisation at each multi-resolution level, and the same groupwise atlas was used in all experiments to enable direct comparison between methods.

4.2 TBM Registrations

Each image was affinely registered with nine degrees of freedom (rigid + scale in each direction) to account for global brain size. The TBM feature data was created by non-rigidly registering each subject to the atlas space. Three methods are compared: One with a fixed regularisation registration strategy, using the same parametrisation as that used to create the groupwise atlas. An approach with a globally adaptive level of bending energy, and the proposed spatially adaptive prior. All the registrations were run to a 10mm knot spacing.

[1] `http://adni.loni.ucla.edu`

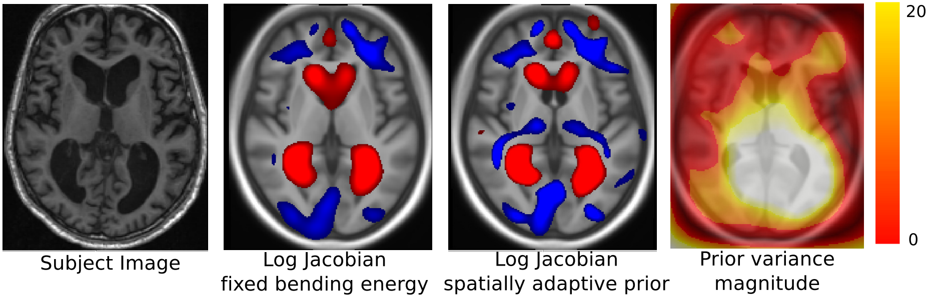

Fig. 1. An example registration from a target subject, who suffers from AD, to the groupwise atlas. The log Jacobian maps show red for regions with expansion greater than 0.5, and blue for contraction. The right hand image shows the magnitude (summed over directions) of the prior variance at each control point overlayed on the atlas.

Once the registrations have been performed, the logarithm of the voxelwise determinant of the Jacobian map is calculated. This provides a measure of local expansion or contraction. These images were input to a general linear model, using age as a regressor, and t-tests were used to evaluate statistical difference between subject groups. The Jacobian maps were not smoothed prior to analysis. All the analysis was performed using tools from the FSL library[2].

5 Results

An example registration using the proposed spatially adaptive prior is shown in Figure 1. As can be seen, using a fixed level of bending energy imposes unnecessary spatial smoothness in the deformation field, and contraction of the thalami is not visible. unlike with the proposed method. Furthermore, the proposed prior provides a lower average deformation magnitude, 2.02mm as opposed to 3.01mm. The map of the prior variance magnitude shows larger deformations are permitted in the region surrounding the ventricles, and the posterior cortical regions.

Figure 2 shows test statistics from the GLM analysis. The use of spatially varying regularisation provides features that are better localised, and therefore more interpretable. Conversely, the use of a global prior enforces a smoothness scale on the obtained deformation features, rather than adapting to the local scale of morphological difference that are indicated by the data.

The proposed regularisation model leads to substantially smaller deformations, as illustrated in Figure 3. As Figure 1 and 2 show that the pathological regions are being deformed, this indicates a reduction in unrelated deformations.

Global regularisation models encourage deformations with global smoothness. The local adaptivity of the proposed method can be illustrated by the standard deviation of the harmonic energy (squared Frobenius norm of the deformation

[2] www.fmrib.ox.ac.uk/fsl/

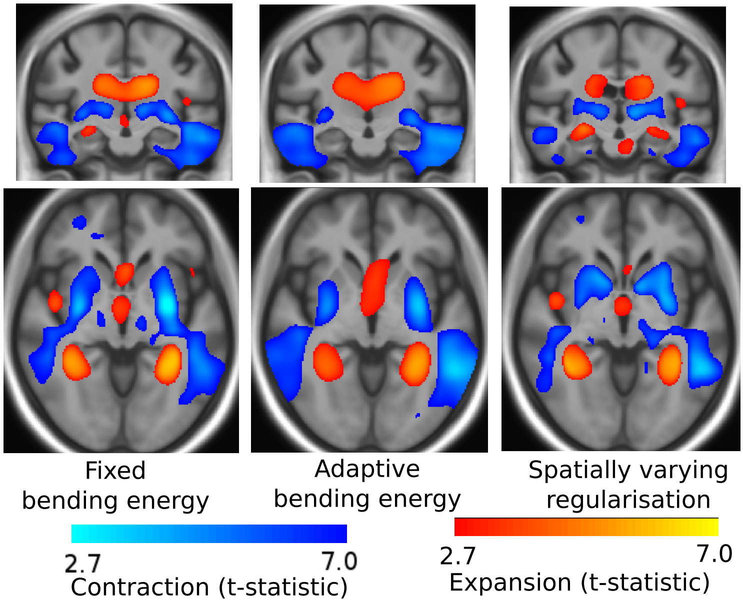

Fig. 2. T-test statistics from the GLM analysis. Red corresponds to expansion in the AD population vs. HC, and blue contraction. The use of bending energy to regularise the registration led to over-smoothed statistics, where the actual region of change is hard to identify. The use of a spatially varing prior better localises the effects. In particular: the expansion of CSF surrounding the hippocampus, and contraction of the thalami and putamen.

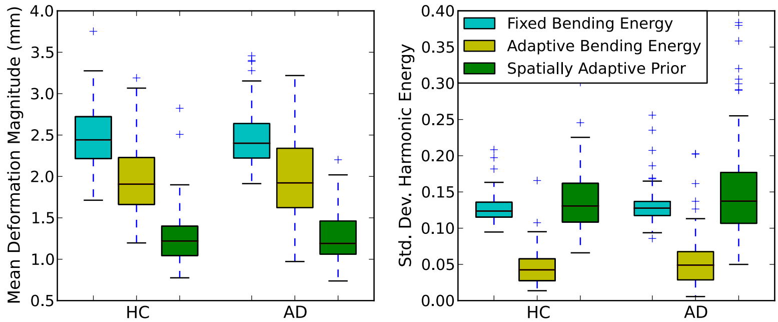

Fig. 3. Boxplot of the average deformation magnitude and the standard deviation of the harmonic energy of the inferred deformation fields across the registrations. The use of a spatially adaptive prior infers much smaller average deformations. The increased standard deviaition of the harmonic energy when using a spatially adaptive prior indicates that the smoothing influence of the regularisation is adaptively varied across the image. The reduction in the strength of the prior in regions with strong information enables the capture of deformations at a data-driven scale.

Jacobian). This metric describes the irregularity of a deformation. In Figure 3 the spatially adaptive prior is demonstrated to allow the smoothness of the deformation greater variability across the image.

6 Discussion and Conclusions

Spatially adaptive regularisation has been demonstrated to provide better localisation of features in tensor based morphometry. The use of a more flexible prior results in deformation fields that have smaller deformations, than those obtained using a global regularisation scheme.

The current implementation requires the numerical inverse of the matrices $\mathbf{\Sigma}$ and $\mathbf{\Upsilon}^{-1}$, which is the computational bottleneck of the approach. This is improved by separating the matrices in terms of deformation directions, which means only an inverse of a symmetric matrix of size $\frac{N_c}{3}$ is required. These inverses could be improved through the use of a Cholesky decomposition, where the inverse could be updated rather than recalculated.

Alternative forms of prior components could be experimented with. This would require no changes to the inference stratergy, and the optimal form of prior components could potentially be evaluated using Bayesian model comparison.

This paper has proposed a model and inference scheme for spatially varying regularisation in registration. This has been shown to improve the localisation of features in TBM, and provides deformation fields where the complexity is adaptively driven by the data. Future work includes extending this model to allow spatially varying noise estimates. The effects of the spatially adaptive prior on registration uncertainty will also be subject to further investigation.

References

1. Ashburner, J., Friston, K.: Voxel-based morphometry–the methods. Neuroimage 11, 805–821 (2000)
2. Lester, H., Arridge, S.R., Jansons, K.M., Lemieux, L., Hajnal, J.V., Oatridge, A.: Non-linear registration with the variable viscosity fluid algorithm. In: Kuba, A., Sámal, M., Todd-Pokropek, A. (eds.) IPMI 1999. LNCS, vol. 1613, pp. 238–251. Springer, Heidelberg (1999)
3. Davatzikos, C.: Spatial transformation and registration of brain images using elastically deformable models. Computer Vision and Image Understanding 66(2), 207–222 (1997)
4. Staring, M., Klein, S., Pluim, J.: Nonrigid registration with tissue-dependent filtering of the deformation field. Physics in Medicine and Biology 52, 6879 (2007)
5. Hermosillo, G., Chefd'Hotel, C., Faugeras, O.: Variational methods for multimodal image matching. International Journal of Computer Vision 50(3), 329–343 (2002)
6. Risholm, P., Samset, E., Wells III, W.: Bayesian estimation of deformation and elastic parameters in non-rigid registration. In: Fischer, B., Dawant, B.M., Lorenz, C. (eds.) WBIR 2010. LNCS, vol. 6204, pp. 104–115. Springer, Heidelberg (2010)

7. Simpson, I., Schnabel, J., Groves, A., Andersson, J., Woolrich, M.: Probabilistic inference of regularisation in non-rigid registration. NeuroImage 59, 2438–2451 (2012)
8. Ashburner, J., Friston, K.: Nonlinear spatial normalization using basis functions. Human Brain Mapping 7(4), 254–266 (1999)
9. Friston, K., Harrison, L., Daunizeau, J., Kiebel, S., Phillips, C., Trujillo-Barreto, N., Henson, R., Flandin, G., Mattout, J.: Multiple sparse priors for the M/EEG inverse problem. NeuroImage 39(3), 1104–1120 (2008)
10. Attias, H.: A variational Bayesian framework for graphical models. Advances in Neural Information Processing Systems 12(1-2), 209–215 (2000)
11. Guimond, A., Meunier, J., Thirion, J.: Average brain models: A convergence study. Computer Vision and Image Understanding 77(2), 192–210 (2000)