Uncertainty Estimates for Improved Accuracy of Registration-Based Segmentation Propagation using Discrete Optimisation

Mattias P. Heinrich[1,2], Ivor J.A. Simpson[3], Mark Jenkinson[2], Sir Michael Brady[4], and Julia A. Schnabel[1]

[1] Institute of Biomedical Engineering, Department of Engineering, University of Oxford, UK
[2] Oxford University Centre for Functional MRI of the Brain, UK
[3] Centre for Medical Image Computing, University College London, UK
[4] Department of Oncology, University of Oxford, UK
mattias.heinrich@eng.ox.ac.uk, http://users.ox.ac.uk/~shil3388

Abstract. A common approach to study developmental-, or pathological differences in the human brain across subjects is to compare differences of the volume and/or shape of anatomical structures. This is often too time consuming to be done manually. Accurate automated methods for brain segmentation, e.g. registration-based segmentation propagation, are therefore important for studying populations. The difficulty and reliability of registering a given annotated atlas to a subject will vary depending on the atlas due to morphological variability. However, few registration algorithms quantify the spatial uncertainty of an obtained transformation or make use of the distribution of uncertainty estimates. In this paper, we describe a novel method of incorporating uncertainty estimates, which are densely evaluated over the space of possible transformations. Based on an efficient discrete optimisation framework, approximately globally optimal marginals for a large range of local displacements are calculated, which can be directly converted into probabilities. We demonstrate that for segmentation propagation in MRI brain scans the use of these probabilities can significantly improve segmentation accuracy. Our method achieves state-of-the-art performance on a publicly available dataset with very low computational cost. Furthermore, for segmentation using multiple atlases, the framework provides an elegant solution to the problem of label fusion.

Keywords: marginal distributions, non-rigid registration, brain labelling

1 Introduction

Deformable image registration, which aims to find anatomical or functional correspondences across scans, is an integral part of numerous applications in medical image analysis. In order to study the functionality, development, and degeneration of human brains, registration-based segmentation propagation is widely used

to automatically label anatomical structures in medical scans [9]. Registration-based segmentation propagation can be used to measure volume and shape of anatomical structures in the human brain. Algorithms using gradient-based optimisation are susceptible to local minima and biased to initialisation. Discrete optimisation techniques offer several advantages over their more commonly used continuous counterparts:

- no derivative of the image similarity metric is required
- their computational efficiency is very high
- the space of permissible deformations can be defined very densely
- an approximately global optimum (instead of a local optimum) can be found

The main limitation of most registration algorithms (regardless of their optimisation) is that only a single optimum defining the most probable transformation is found. Quantifying the uncertainty distribution of a registration enables us to consider all probable deformations rather than only the strongest optimum.

We present a new approach, which determines very good approximations to the marginal distributions of all potential transformations. The deformations are parameterised by a very densely quantised space of displacements. We show that the propagation of segmentation labels can be significantly improved by making use of this rich information of uncertainty. Additionally, it yields a direct measurement of the local reliability of the segmentation, which can be very valuable in many applications.

In the next section, we review a number of recent approaches, which make use of registration uncertainties in different contexts. In Sec. 3, we present the image registration model and the details of the optimisation. In Sec. 4, we explain how uncertainty estimates can be used for segmentation propagation using a weighted voting. The method is evaluated and validated in Sec. 5 for for the automatic labelling of 56 cortical structures in magnetic resonance imaging (MRI) scans.

2 Related Work

Probabilistic registration methods based on continuous optimisation have been used to estimate the spatial variation of the displacements close to a local optimum to improve deformable registration based on locally adaptive smoothing [17, 18]. In [8] uncertainty of registration parameters is used to improve segmentation propagation by using multiple probable warps from atlas to target volume. Registration uncertainties are also used in [14] to estimate the cumulated dose delivery in radiotherapy and in [13] to estimate registration accuracy. The limitation of these approaches, based either on bootstrapping [13], variational Bayes [17, 18] or Monte Carlo sampling [8, 14], is that a dense sampling of the uncertainty of the displacement space is impossible and distributions with multiple local optima cannot be dealt with easily.

Previous work in discrete optimisation has mainly focussed on the estimation of approximate marginals using dynamic FastPD, a graph cut optimisation [12]. A dynamic refinement of the displacement space has been proposed in [4] for

optical flow estimation. Here a covariance matrix of a Gaussian distribution was fitted to the distribution and used to restrict the spatial sampling range of displacements for the next iteration. Obtaining min-marginal distributions using graph cuts is possible [10, 19], however the optimality guarantee can become poor for very large label spaces (i.e. in 3D medical image registration).

In the next section, we will present the employed registration model. This will be followed by a description of the proposed optimisation scheme, which estimates dense marginal distributions over large displacement spaces using belief propagation (BP) on a simplified graph structure (a spanning tree). For this relaxed graph structure, BP enables the computation of exact, globally optimal marginals. Note that the related approach of dynamic programming, used e.g. in [3, 20], only finds the most probable displacements.

3 Registration Model and Optimisation

Parametric deformable image registration using discrete optimisation can be formulated on a graph structure, in which nodes $p \in \mathcal{P}$ (with coordinates $\mathbf{x}_p$) correspond to control points of the transformation grid [5]. The space of potential displacements $\mathbf{u}_p = \{u_p, v_p, w_p\}$ for each control point is discretised with a quantisation step of d yielding a three dimensional displacement space of $\mathcal{L} = \{0, \pm d, \pm 2d, \ldots, \pm \max(|\mathbf{u}|)\}^3$. The image similarity $\mathcal{S}$ between target scan I and moving scan J can be evaluated independently for each node by taking into account all voxels in spatial proximity $\eta(\mathbf{x}_p)$ and assuming a constant displacement vector for them. Since no derivative of the similarity term is required, any point-wise metric can be used. In this work, we use absolute differences of intensity gradients, which are less affected by locally changing image contrast than sums of squared differences (SSD):

$$\mathcal{S}(\mathbf{x}_p, \mathbf{u}_p) = \sum_{\mathbf{x}_i \in \eta(\mathbf{x}_p)} |\nabla I(\mathbf{x}_i) - \nabla J(\mathbf{x}_i + \mathbf{u}_p)| \tag{1}$$

Smooth transformations can be ensured by avoiding large differences in displacement vectors for neighbouring nodes with a regularisation penalty $\mathcal{R}$. In this work, absolute differences (known as total variation regularisation) are used. For each edge in the graph, which connects two nodes ($(p, q) \in \mathcal{N}$), a penalty is incurred for differences of pair-wise displacements:

$$\mathcal{R}(\mathbf{u}_p, \mathbf{u}_q) = \sum_{(p,q) \in \mathcal{N}} \frac{|\mathbf{u}_p - \mathbf{u}_q|}{|\mathbf{x}_p - \mathbf{x}_q|} \tag{2}$$

A weighting parameter λ sets the influence of the regularisation in a combined energy function:

$$E(\mathbf{u}) = \sum_{p \in \mathcal{P}} \mathcal{S}(\mathbf{x}_p, \mathbf{u}_p) + \lambda \mathcal{R}(\mathbf{u}_p, \mathbf{u}_q) \tag{3}$$

Minimising this energy for arbitrary graph structures is NP-hard. One solution to solve Eq. 3 is to use sequential tree-reweighted message passing (TRW-S)

Algorithm 1: Estimation of marginals using belief propagation.

1. Initialise marginals and messages:
foreach node **do** marginals [node] $\leftarrow \mathcal{S}(\mathbf{u}_p)$; message [node] $\leftarrow 0$;
2. Forward-pass of messages:
for node = *leaves* **to** *root-1* **do**
  cost $\leftarrow$ marginals [node];
  message [node] $\leftarrow$ `min-sum(cost)`; (see Eq. 4)
  marginals [parent] $\leftarrow$ marginals [parent] + message [node];
end
3. Backward-pass of messages:
for node = *root-1* **to** *leaves* **do**
  cost $\leftarrow$ marginals [parent] - ~~marginals~~ [node] + message [parent];
  message [node] $\leftarrow$ `min-sum(cost)`; (see Eq. 4)
end
4. Add messages to marginals:
foreach node **do** marginals [node] $\leftarrow$ marginals [node] + message [node];

[11]. Here, the problem is simplified by passing messages along individual chains of the graph (which contain no loops and can be optimally solved for). The marginals are averaged after message passing in order to link the three chains passing through every node. The algorithm is run for several iterations to ensure that information can be spread across the graph. Disadvantages are the high computational costs, since each iteration requires the calculation of six message updates per node, and it is not possible to directly parallelise TRW-S due to its sequential nature.

We propose a different approach, which attempts to improve on this concept by finding the exact marginals for each displacement for each node (and the global optimum of E) for the relaxed graph structure of a spanning tree [3, 20]. This concept has been successfully applied in our previous work on deformable lung registration [6, 7]. First, a spanning tree is constructed, which connects all nodes in the graph without loops. At each node p, a message vector $\mathbf{m}_p$ containing the cost for the best displacement $\mathbf{u}_p^*$, given the displacement $\mathbf{u}_q$ of its parent node q and the messages of its children c, can be found by evaluating:

$$\mathbf{m}_p(\mathbf{u}_q) = \min_{\mathbf{u}_p} \left(\mathcal{S}(\mathbf{x}_p, \mathbf{u}_p) + \lambda \mathcal{R}(\mathbf{u}_p, \mathbf{u}_q) + \sum_c \mathbf{m}_c(\mathbf{u}_p) \right) \tag{4}$$

For any leaf node in the tree, Eq. 4 can be evaluated directly (since it has no children). Thereafter, the tree is traversed from its leaves down to the root node. The message vector of (only) the root node then contains the exact marginals, and the best displacements can be found by replacing min with argmin in Eq. 4 by traversing the tree from root to leaves. This is used in classical registration approaches, which only find the most probable transformation.

The marginal distribution for each node can be obtained by another iteration of the min-sum message passing, but this time in the opposite direction. Algorithm 1 shows the pseudo-code to obtain min-marginal distributions. A randomly

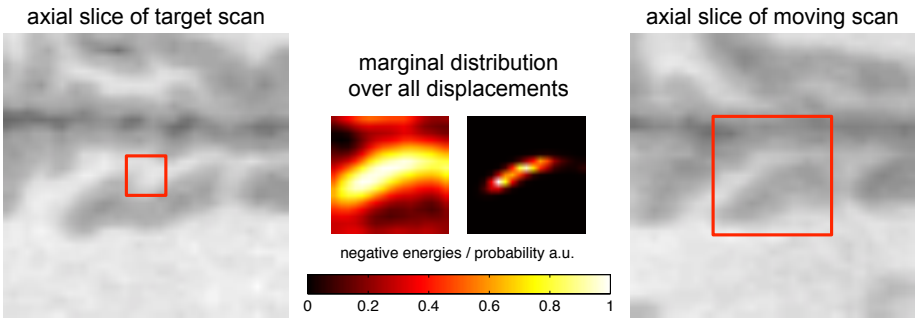


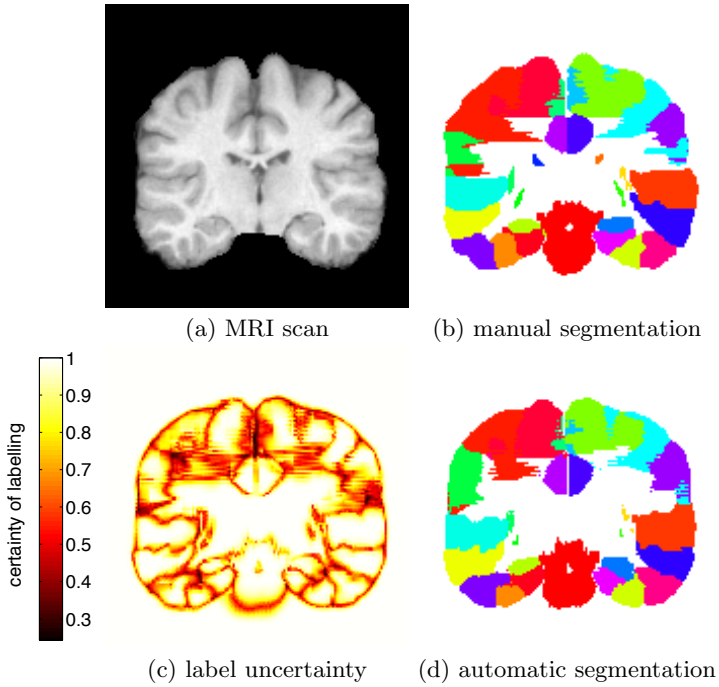
Fig. 1. Concept of the use of uncertainty from marginals in registration. Axial slices of moving and target scans are shown (red box in moving scan indicates the extent of the displacement space $\mathcal{L}$). A dense distribution of marginal energies is obtained using message passing (see Sec. 3) and converted into probabilities using Eq. 5. This yields a sharper distribution (yet still with a certain spatial variation), which can be directly used for a weighted segmentation propagation.

selected spanning tree does not capture all dependencies between nodes. For this reason, multiple trees are chosen and inference is repeated independently for each tree, and the resulting marginal distributions are averaged for each node afterwards. We choose a simple averaging of marginals from different trees, in order to give no prior preference to any single tree. When using a sufficient number of trees the reduced connectedness of each spanning tree can be compensated for, given their variability adequately captures the real neighbourhood structure. We found five trees were enough to achieve this in our experiments. The advantage of using message passing, compared to graph cut approaches is that marginal distributions for each node can be obtained directly, which can then be used to quantify the local uncertainty of the registration (for a proof, see [10]).

Evaluating Eq. 4 naïvely would require $|\mathcal{L}|^2$ calculations, due to the pairwise regularisation cost, for each node. This would prohibit the use of large displacement spaces. In [2] the lower envelope technique is introduced, which can be employed for several commonly used pair-wise penalties (e.g. total variation or diffusion regularisation) and reduces the complexity to $|\mathcal{L}|$. The next section describes, how the marginal distributions can be employed for improved segmentation propagation.

4 Uncertainty Estimates for Segmentation Propagation

To date, few attempts have been made to directly employ the benefits of uncertainty estimates from discrete optimisation, and min-marginals have not been used in medical image registration. This is partly due to the high computational cost of dense displacement spaces of graph cuts, the high memory demand of FastPD, and the weak convergence of loopy BP [19]. Here, the use of BP on a



(a) MRI scan (b) manual segmentation

(c) label uncertainty (d) automatic segmentation

Fig. 2. (a) Coronal slice of target MRI scan. (b) Manual segmentation of target scan. (c) Local certainty of label propagation (high uncertainty indicated by dark colours). (d) Propagated segmentations from atlas scan using proposed method. At interfaces between segmentations the uncertainty is highest. Here, the weights could be used to estimate partial tissue volume. Note that the manual segmentation were performed on axial slices. Therefore, some artefacts occur at label boundaries in both manual and automatic segmentation, even though all processing steps are performed in 3D.

tree improves the accuracy of the marginals approximation compared to loopy BP and greatly reduces memory demand and computation time with respect to graph cut solutions.

Several challenges are common in inter-subject registration, which can cause the occurrence of local uncertainties of the transformations:

- missing one-to-one correspondences of anatomies across subjects
- mismatch of structures due to local minima of the registration
- untrustworthy image information due to acquisition noise and artefacts

In the following, we show how a probabilistic label propagation approach can be used to compensate (in parts) for inaccuracies introduced by these challenges.

Probabilistic label propagation: For the propagation of segmentations from a manually labelled atlas, estimating a transformation between a target and moving scan is only an intermediate goal. In fact, choosing only the most

probable of many possible transformations can limit the accuracy of the obtained segmentation. The estimated marginal distributions enable the use of many different probable mappings, which can be weighted by their probability. Note, that in [8] many probable transformations were sampled using Monte Carlo sampling. However, no fractional weighting could be obtained. The probability $p(\mathbf{x}_i, \mathbf{u}_i)$ for each voxel $i \in \Omega$ in the image and each displacement $\mathbf{u}_i \in \mathcal{L}$ can be directly obtained from the min-marginal energies $E(\mathbf{x}, \mathbf{u})$[10] when normalised by n:

$$p(\mathbf{x}_i, \mathbf{u}_i) = \frac{1}{n} \exp\left(-\frac{\beta \cdot E(\mathbf{x}_i, \mathbf{u}_i)}{\mathrm{std}(E(\mathbf{x}_j, \mathbf{u}_j))}\right) \quad \mathbf{u}_j \in \mathcal{L} \text{ and } \mathbf{x}_j \in \Omega \tag{5}$$

Dividing each marginal energy by the standard deviation over all marginals compensates for a global offset of E. β varies the spread of the probability estimates: low values result in smoother distributions, and high values in narrower peaks ($\beta \to \infty$ is equivalent to choosing only the transform with lowest energy). The choice of this parameter will be discussed later. Finding the optimal segmentation label is now possible by summing the probabilities for each possible segmentation label (from the atlas) and choosing the arg max of all labels. Due to memory limitations, a parametric transformation model is chosen with a control point spacing of more than one voxel. To define the segmentation label for each voxel, the marginal distributions are interpolated by trilinear interpolation between control points, resembling a first order B-spline transformation model.

Symmetric label probability estimation: Deformable registration should in general be treated as a symmetric problem [1]. When the marginals for both forward and backward transformation are estimated, the probabilities can easily be combined for each voxel and segmentation label. For the classical backward transformation (unlabelled scan is target), the displacement space for each voxel may cover different possible segmentation labels. For the forward transformation for each voxel only one segmentation label is propagated into the target space (where the probabilities from different propagations are accumulated).

Fusion of multiple atlases: Much research has addressed finding a good fusion of labels from multiple atlases, see e.g. [15]. In our approach incorporating multiple atlases is straightforward once the marginals are estimated between the target scan and each labelled moving scan. The votes from each atlas are aggregated for each voxel and the segmentation label with highest probability is chosen thereafter. By using the uncertainty estimates a higher weighting of the "best" atlas is intrinsically given, making this approach fully adaptive to the uncertainty in different anatomical regions.

5 Experimental Evaluation of Segmentation Propagation for MRI Brain Scans

Experiments on label propagation for automatic segmentation of human brain scans were performed using the LBPA40 dataset [16]. This dataset includes MRI scans of 40 normal adults and manual segmentations of 56 anatomical structures.

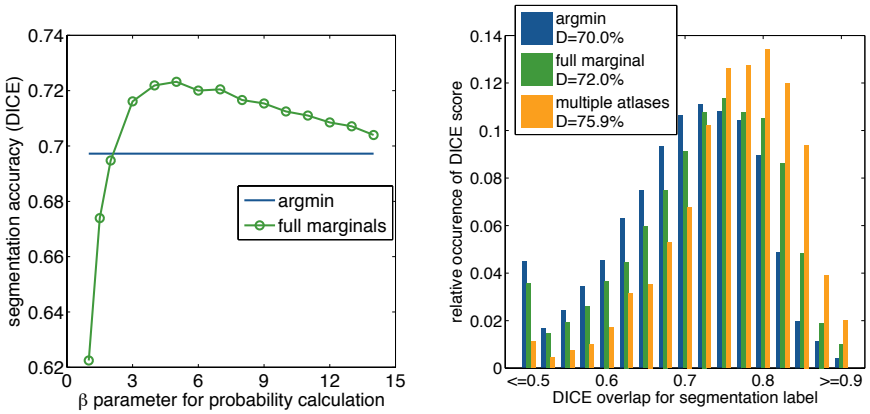


Fig. 3. Left: Influence of β for the conversion of min-marginals to probabilities in Eq. 5 ($\beta \to \infty$ is equivalent to the argmin). Right: Distribution of DICE scores for experiments. Improvements of 2% are obtained when using the full marginal distribution instead of taking the argmin. Using three atlases further improves the overlap by 4%.

It has been used in a comprehensive evaluation study of 14 non-linear registration algorithms [9], allowing for a direct quantitative comparison with state-of-the-art methods. We follow the same protocol as [9] and use skull-stripped scans, which are resampled to an isotropic resolution of 1 mm and rigidly aligned to the MNI305 atlas (no further linear alignment was performed in our experiments). We use the first 10 subjects just for parameter tuning. The remaining 30 subjects are divided into three groups. For each subset of 10 scans, we perform one-to-all-others registrations (yielding a total of 270 registration experiments). Figure 2 shows an example of the presented segmentation propagation, together with the obtained uncertainty map. Higher uncertainty is located at the boundaries of segmentation labels, which is partly due to the partial volume effect (and inconsistencies of the manual segmentations, which are labelled in axial planes).

Parameter settings: The range of displacements ($\max(|\mathbf{u}|)$) was defined to be 8 mm in each direction and dimension, with a quantisation of 2 mm. A control-point grid with spacing of 5 mm was used, yielding roughly $2.5{\times}10^7$ degrees of freedom. The threshold for intensity gradients was set to 50, which is around a quarter of the scans mean intensity. $\lambda = 50$ was empirically chosen for the weighting of the regularisation, which gave the best segmentation accuracy for both variants of our algorithm (doubling or halving λ reduces the accuracy by less than 0.3%). Figure 3 (left) shows the influence of the β parameter of Eq. 5. It can be seen that too small values of β cause an over-smoothing of the probability distributions, larger values are similar to the classical approach, and best accuracies are obtained for $\beta \approx 5$.

Quantitative Results and Comparisons: Segmentation accuracy is measured with the DICE coefficient, which is defined as: $\mathrm{D}{=}(2|T \cap M|)/(|T| + |M|)$,

where T and M are the segmentation labels in target and moving scan. Using the proposed probabilistic label weighting scheme, the segmentation accuracy was improved by 2% to 72.0% (see Fig. 3 right). The computation time for each registration is less than 60 seconds on a single CPU core. We also compared our optimisation strategy to TRW-S, using five iterations of message-passing, which achieves almost identical results (+0.07%), albeit with a three times higher computational burden. Our results are similar to **syn** [1], which performed best in the study of [9] and achieves a target overlap of 71.5% (our result for target overlap is 71.4%), but is computationally much more expensive (time per registration is about one hour). The improvements of using the full-marginal distribution compared to the standard argmin approach are significant for 53/56 labels ($p < 0.05$) and occur most commonly at locations of anatomical shapes, which pose a higher difficulty for the standard registration approach (e.g. due to a high curvature). Further improvements can be achieved when using multiple atlases for each label propagation (D= 75.9%, for three atlases).

6 Discussion and Conclusion

A novel framework for incorporating uncertainty estimates of displacement parameters to improve segmentation propagation is presented. An efficient discrete optimisation strategy, using BP on random spanning trees, is used to obtain dense estimates of the marginal distribution of possible transformations, in contrast to only parametric estimates used in previous work (e.g. [18]). Our method shows significant (for 53/56 labels) improvements over the standard approach (on average +2%), which only takes the most probable transformation into account. The approach is validated by segmentation propagation experiments on MRI brain scans of 40 adults with 56 manually labelled anatomical structures. The obtained volume overlap of 72.0% is comparable to the best state-of-the-art methods with more than a ten fold reduced computation time. Further improvements could be made for multi-atlas segmentation, by using the cost of the maximum a posterior solution to discard scans, which failed to be registered.

Future work will include the application of our method to the MICCAI SATA segmentation challenge, to facilitate a more detailed comparison to other state-of-the-art (multi)-atlas segmentation methods. Furthermore, the impact of an improved segmentation on the classification accuracy of neurodegenerative disease should be studied, as e.g. done in [8]. The use of the estimated displacement probability distributions is not limited to segmentation propagation. For the estimation of non-rigid motion from cardiac or respiratory sequences, uncertainty estimates could be used to improve subvoxel accuracy.

Acknowledgements: We would like to thank EPSRC and Cancer Research UK for funding this work within the Oxford Cancer Imaging Centre.

References

1. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly

and neurodegenerative brain. Medical Image Analysis, 12(1) pp. 26–41 (2008)
2. Felzenszwalb, P., Huttenlocher, D.: Efficient Belief Propagation for Early Vision. International Journal of Computer Vision 70(1), 41–54 (2006)
3. Felzenszwalb, P., Huttenlocher, D.: Pictorial structures for object recognition International Journal of Computer Vision 61(1), 55–79 (2005)
4. Glocker, B., Paragios, N., Komodakis, N., Tziritas, G., Navab, N. Optical Flow Estimation with Uncertainties through Dynamic MRFs. IEEE CVPR (2008)
5. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registration through MRFs and efficient linear programming. Medical Image Analysis 12(6), pp. 731–741 (2008)
6. Heinrich, M.P., Jenkinson, M., Brady, M., Schnabel, J.A.: Globally Optimal Deformable Registration on a Minimum Spanning Tree using Dense Displacement Sampling. In: Ayache, N., Delingette, H., Golland, P., Mori K., MICCAI 2012, LNCS, vol. 7512, pp. 115–122, (2012)
7. Heinrich, M.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), pp. 1239–1248 (2013)
8. Iglesias, J.E., Sabuncu, M.R., van Leemput, K.: Improved Inference in Bayesian Segmentation Using Monte Carlo Sampling: Application to Hippocampal Subfield Volumetry. Medical Image Analysis 17(7), pp.766–778, (2013)
9. Klein, A. et al: Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage 46(3), 786–802 (2008)
10. Kohli, P., Torr, P.H.S.: Measuring uncertainty in graph cut solutions. Computer Vision and Image Understanding 112(1), 30–38 (2008)
11. Kolomogorov, V.: Convergent tree-reweighted message passing for energy minimization. IEEE Trans. Pattern Anal. Mach. Intell. 28(10), 1568–1583 (2006)
12. Komodakis, N., Tziritas, G., Paragios, N.: Performance vs computational efficiency for optimizing single and dynamic MRFs: Setting the state of the art with primal-dual strategies. Computer Vision and Image Understanding 112(1), 14–29 (2008)
13. Kybic, J.: Bootstrap resampling for image registration uncertainty estimation without ground truth. IEEE Transactions on Image Processing 19(1) 64–72 (2010)
14. Risholm, P., Balter, J., Wells, W.M.: Estimation of Delivered Dose in Radiotherapy: The Influence of Registration Uncertainty In: Fichtinger, G., Martel, A. and Peters, T., MICCAI 2011, LNCS, vol. 6892, pp. 548–555, (2011)
15. Sabuncu, M.R., Yeo, B.T.T., van Leemput, K., Fischl, B., Golland, P.: A generative model for image segmentation based on label fusion. IEEE Transactions on Medical Imaging 29(10) 1714–1729 (2010)
16. Shattuck, D.W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K.L., Poldrack, R.A., Bilder, R.M., Toga, A.W. Construction of a 3D probabilistic atlas of human cortical structures. NeuroImage 39, 1064–1080 (2008)
17. Simpson, I.J.A, Woolrich, M.W., Groves, A.R., Schnabel, J.A. Longitudinal Brain MRI Analysis with Uncertain Registration In: Fichtinger, G., Martel, A. and Peters, T., MICCAI 2011, LNCS, vol. 6892, pp. 647–654, (2011)
18. Simpson, I.J.A., Woolrich, M.W., Schnabel, J.A. Probabilistic segmentation propagation from uncertainty in registration Proceedings of Medical Image Understanding and Analysis (MIUA) (2011)
19. Tarlow, D., Adams, R.P.: Revisiting uncertainty in graph cut solutions IEEE CVPR, pp. 2440-2447, (2012)
20. Veksler, O.: Stereo Correspondence by Dynamic Programming on a Tree. IEEE CVPR 2(2) pp. 384–390 (2005)