

Deformable Image Registration by Combining Uncertainty Estimates from Supervoxel Belief Propagation

Mattias P. Heinrich^{a,*}, Ivor J.A. Simpson^b, Bartłomiej W. Papież^c,
Sir Michael Brady^d, Julia A. Schnabel^c

^a*Institute of Medical Informatics, Universität zu Lübeck, Germany*

^b*Centre for Medical Image Computing, University College London, UK*

^c*Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, UK*

^d*Department of Oncology, University of Oxford, UK*

Abstract

Discrete optimisation strategies have a number of advantages over their continuous counterparts for deformable registration of medical images. For example: it is not necessary to compute derivatives of the similarity term; dense sampling of the search space reduces the risk of becoming trapped in local optima; and (in principle) an optimum can be found without resorting to iterative coarse-to-fine warping strategies. However, the large complexity of high-dimensional medical data renders a direct voxel-wise estimation of deformation vectors impractical. For this reason, previous work on medical image registration using graphical models has largely relied on using a parameterised deformation model and on the use of iterative coarse-to-fine optimisation schemes. In this paper, we propose an approach that enables accurate voxel-wise deformable registration of high-resolution 3D images without the need for intermediate image warping or a multi-resolution scheme. This is achieved by representing the image domain as multiple comprehensive supervoxel layers and making use of the full marginal distribution of all probable displacement vectors after inferring regularity of the deformations using belief propagation. The optimisation acts on the coarse scale representation of supervoxels, which provides sufficient spatial context and is robust to noise in low contrast areas. Minimum spanning trees, which connect neighbouring supervoxels, are employed to model pair-wise deformation dependencies. The optimal displacement for each voxel is calculated by considering the probabilities for all displacements over all overlapping supervoxel graphs and subsequently seeking the mode of this distribution. We demonstrate the applicability of this concept for two challenging applications: first, for intra-patient motion estimation in lung CT scans; and second, for atlas-based segmentation propagation of MRI brain scans. For lung registration, the voxel-wise mode of displacements is found using the mean-shift algorithm, which enables us to determine continuous valued sub-voxel motion vectors. Finding the mode of brain segmentation labels is performed using a voxel-wise majority voting weighted by the displacement uncertainty estimates. Our experimental results show significant improvements in registration accuracy when using the additional information provided by the registration uncertainty estimates. The multi-layer approach enables fusion of multiple complementary proposals, extending the popular fusion approaches from multi-image registration to probabilistic one-to-one image registration.

Keywords: supervoxel layers, segmentation propagation, motion estimation, mean-shift, probabilistic registration

1. Introduction

Medical image registration aims to find spatial correspondences between scans of different patients, modalities, the time course of a disease, or response to therapy. It also forms an integral part of many medical image analysis applications. For example, intra-patient deformable registration can be used to relate two scans, e.g. a pre-treatment planning scan to an intra-operative image for guiding an intervention. Longitudinal scans can be employed to monitor treatment or disease progression. Computed tomography (CT) scans are now used widely for motion estimation in radiotherapy planning in order

to increase the accuracy of dose delivery (Weiss et al., 2007). To study the functionality or anatomical variability of human brains, registration-based segmentation propagation is widely used to automatically label structures in magnetic resonance images (MRI) (Klein et al., 2009). It can be used to measure the volume and shape of anatomical structures in the human brain.

Image registration algorithms are generally based on three components: a similarity metric, a transformation model, and an optimisation strategy. A large variety of approaches have been proposed for the medical image domain over the past few years (see (Sotiras et al., 2013) for an overview). Different assumptions have been made to model these three components to obtain robust, accurate and also computationally efficient algorithms. In the following, we discuss four challenges, which are, in our opinion, prevalent in current approaches.

^{*}Corresponding author:

Email address: heinrich@imi.uni-luebeck.de (Mattias P. Heinrich)

URL: <http://www.mpheinrich.de> (Mattias P. Heinrich)

First, the similarity metric measures the data affinity and can assume many different forms depending on the medical application and modality. Ideally, the choice of the similarity measure is affected by neither the chosen transformation model nor the employed optimisation technique. In practice, however, this is often not the case, since gradient-based optimisation techniques require the similarity term to be (at least first-order) differentiable.

Second, the transformation model may restrict the deformation between two images either to obey a certain physically motivated model (for example finite element models (FEM) (Ferrant et al., 2000)) or to be well approximated by a particular mathematical model (such as free-form transformations (Rueckert et al., 1999)). Due to its simplicity, most common registration algorithms define a transformation model on a regular, equally spaced grid. However, sparsely or irregularly spaced models, which have been e.g. presented by (Schnabel et al., 2001), (Rohde et al., 2003) (with spline basis functions), (Glocker et al., 2010) (triangular mesh) and (Popuri et al., 2013) (FEM), might in many cases provide a more realistic model of real physical motion, in particular when the smoothness of deformations varies across the image domain.

Third, the chosen optimisation strategy impacts the space of obtainable deformation vectors. Continuous optimisation approaches yield excellent results for subtle sub-voxel changes across scans, which is important for the analysis of longitudinal brain development (Ashburner and Ridgway, 2012). A disadvantage is that they are susceptible to local minima, especially in the presence of large, complex motions.

Fourth, a further limitation of the majority of current methods is that they only estimate the most probable transformation (the *maximum a posteriori solution*). However, quantifying the uncertainty distribution of a registration can provide improvements with respect to the immediate goals, such as segmentation or motion vector estimation, as well as give a confidence measure of the generated results.

Recent work (Shekhovtsov et al., 2008; Glocker et al., 2008a; Heinrich et al., 2013a; Cobzas and Sen, 2011), has demonstrated a number of advantages of discrete optimisation techniques over the more commonly used continuous counterparts: they do not necessitate the computation of derivatives of the image similarity metric; they are computationally efficient; the space of displacements can be defined to capture a large range of deformations; and, under certain circumstances, local minima can be avoided. For a more in-depth discussion on optimality guarantees see (Komodakis and Tziritas, 2007) for linear programming and (Felzenszwalb and Zabih, 2011) for belief propagation techniques. Discrete approaches are, however, restricted to a quantisation of displacement vectors, which cause limits on the achievable accuracy. An extended review of deformable medical image registration using Markov random field formulations can be found in (Glocker et al., 2011).

In this paper, we further contribute to medical image registration based on graphical models by introducing three new concepts, which have to date not been deeply explored. First, we make use of a more flexible image representation using supervoxel graphs. This is, to the best of our knowledge, the first

time that multiple complementary layers have been used, which enable us to represent the complex nature of 3D deformations with spatially varying smoothness. Second, the probability (inversely, uncertainty) of a large set of potential displacements is calculated for our graphical model using belief propagation (with the min-sum algorithm). Supervoxels are inter-connected using a tree, which enables us to obtain marginal energies for every displacement and to regularise the displacement field using pair-wise interactions. Third, the complementary information from multiple supervoxel layers is combined on a voxel-wise level. A mode seeking algorithm over all potential (and probable) displacements for every voxel is used to find subvoxel accurate motion vectors (or the most likely fusion of many potential segmentation labels).

This paper builds upon previous work by the authors including the use of a graphical model that represents the image domain by a number of overlapping layers of supervoxels, which are connected by a minimum spanning tree (MST) (Heinrich et al., 2013b); the optimisation of the MST model using belief propagation (Heinrich et al., 2013a) and (Heinrich et al., 2012) and a similar calculation of min-marginal probabilities over all potential displacements (Heinrich et al., 2013d). Here, we unify those approaches, extend the motion estimation by the mean-shift mode seeking algorithm, and report substantial additional experimental validation.

2. Background

Supervoxel: Superpixel clustering describes the parcellation of the image domain into perceptually meaningful regions, which group pixels based on their appearance and spatial closeness. (Ren and Malik, 2003) introduced the term superpixel, but previous work, such as watershed segmentation, have followed the same principle. Because they significantly reduce computational complexity, they have attracted a lot of attention in a range of image analysis tasks including optical flow estimation (Lei and Yang, 2009; Zitnick and Kang, 2007); and image segmentation (Shi and Malik, 2000). In medical imaging, 3D supervoxels have been introduced relatively recently for cell segmentation by (Lucchi et al., 2012). Brain tumour segmentation has been addressed by (Wang and Yushkevich, 2013) using supervoxel matching without regularisation, following the superparsing framework of (Tighe and Lazebnik, 2013). Another recent approach by (Tang and Hamarneh, 2014) used supervoxels in an aggregation step for random walk registration. In (Felzenszwalb and Huttenlocher, 2006) a uniform hierarchical grouping of nodes was used to improve the convergence of belief propagation, while (Willsky, 2002) proposed organised trees of multiple scales in a pyramid form to solve large scale MRF problems.

Uncertainty estimates: 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 (Simpson et al., 2011) and boost classification accuracy (Simpson et al., 2013). This Bayesian framework was extended by Wassermann et al. (2014) for large diffeomorphic

mappings. Iglesias et al. (2013) use uncertainty of registration parameters to improve segmentation propagation by using multiple probable warps from atlas to target volume. Registration uncertainties are also used in (Risholm et al., 2013) to estimate the cumulated dose delivery in radiotherapy and in (Kybic, 2010) to estimate registration accuracy. The limitation of these approaches, based either on bootstrapping (Kybic, 2010), variational Bayes (Simpson et al., 2012), or Monte Carlo sampling (Iglesias et al., 2013; Risholm et al., 2013), is that a dense sampling of the uncertainty of the displacement space is impossible (or at least extremely computationally expensive) and distributions with multiple local optima cannot be easily dealt with. The AQUIRC method (Datterri and Dawant, 2012) has been used to estimate registration uncertainty (and thus accuracy) using transitivity error in registration circuits (at least four pair-wise registrations) and combine the best of multiple atlas-based registrations to reduce the localisation error of anatomical landmarks.

(Rohde et al., 2003) and (Schnabel et al., 2001) use the local registration uncertainty during an iterative approach to determine an adaptive set of active control points / or image regions. Both approaches rely on image similarity information to determine areas that are not yet correctly aligned. (Oreshkin and Arbel, 2013) have introduced a Bayesian framework, which enables a substantial subsampling of voxels used in the optimisation of rigid transforms.

Previous work in discrete optimisation has mainly focused on the estimation of approximate marginals using dynamic FastPD, a graph cut optimisation (Komodakis and Tziritas, 2007). A dynamic refinement of the displacement space has been proposed in (Glocker et al., 2008b) for optical flow estimation. They fit a multivariate Gaussian distribution to the displacement distribution and use the covariance matrix to restrict the spatial sampling range for the next iteration. This was later applied to concurrent medical image registration and segmentation (Parisot et al., 2014), where, additionally, the sparseness of the underlying graph nodes was determined by registration uncertainty. Obtaining min-marginal distributions using graph cuts is possible (Kohli and Torr, 2008; Tarlow and Adams, 2012), however the optimality guarantee can become poor for very large label spaces (i.e. in 3D medical image registration). In (Tang and Hamarneh, 2013) the displacement space is iteratively increased based on the probabilistic cost of a coarser label space within a random walk registration framework. A limitation, compared to our approach, is that the set of labels has to be the same for all control points.

Combination and fusion of label proposals: Fusion of propagated segmentations from multiple atlases has been widely studied (Warfield et al., 2004), especially for the application of brain labelling. A non-local refinement search around the location of an initial registration (linear or deformable) has been found useful in applications such as patch-similarity based segmentation (Coupé et al., 2011; Rousseau et al., 2011), Similarity and Truth Estimation for Propagated Segmentations (STEPS) (Cardoso et al., 2013) or non-local STAPLE (Asman and Landman, 2013). These non-local approaches have in common that the refinement search is performed without consid-

ering the spatial regularity of displacements (which may have to be corrected for afterwards, c.f. (Wang et al., 2013)). In this work, we unify both registration and label fusion within the same optimisation framework. Furthermore, our approach is well suited to segment an unseen image using a single atlas only.

Fusion of many potential solutions from different algorithms for optical flow estimation has been proposed in (Lempitsky et al., 2008) using a new fusion move for graph cut optimisation. Muenzing et al. (2012b) proposed combining the outcome of several medical image registration techniques applied to lung CT based on a trained local registration quality assessment (Muenzing et al., 2012a). In related work using image-based features, (Forsberg et al., 2011) estimates weights for multichannel images and (Lotfi et al., 2013) learn registration uncertainty with random regression forests. The (near) optimal combination of all algorithms within a boosting framework has been shown to improve registration accuracy for challenging datasets (Muenzing et al., 2014). Yushkevich et al. (2012) used a generalisation of their segmentation fusion approach (Wang et al., 2013) to combine several correspondences (deformation fields) in a group wise registration setting. Multiple runs with varying smoothness parameters of the demons algorithm have been used in (Ye et al., 2013) to find the optimal local regularisation weighting and combine the registrations based on this.

3. Methods

The registration method we have developed consists of three steps. First, the generation of multiple supervoxel graphs. Second, the optimisation of a registration cost function and calculation of marginal probabilities for potential displacement vectors. Finally, the combination of the uncertainty estimates of overlapping supervoxels for each single voxel using a mode seeking algorithm.

In Sec. 3.1, the image domain, Ω , is subdivided into a number of overlapping supervoxels, $v \in \mathcal{V}$. The initialisation of the supervoxel clustering algorithm contains a certain degree of variability, therefore the shapes of clusters are slightly different in each layer, and every voxel, \mathbf{x} , is part of a set of differently overlapping supervoxels. Each layer is defined as a graph, \mathcal{G} , consisting of $|\mathcal{V}|$ vertices (supervoxels) that are connected by edges, $e \in \mathcal{E}$, over which the spatial regularity of the deformations is encouraged. The edges in this graph are chosen so that it contains no loops, forming a spanning tree with minimal total edge cost (MST).

Section 3.2 describes the registration method. A discrete displacement space, $\mathbf{d} \in \mathcal{L}$, is defined, which parameterises a displacement field, \mathbf{u} . A cost function, $E(\mathbf{u})$, consisting of an image similarity term, \mathcal{S} , and a regularisation penalty, \mathcal{R} , is defined for each graph. For the case of a graph without loops (i.e. a tree), the globally optimal displacement can be found by using a single pass of belief propagation (also called dynamic programming or the Viterbi algorithm) (Felzenszwalb and Zabih, 2011). A second pass of belief propagation is employed to find the exact min-marginal energies for every possible displacement for every node in the tree. These min-marginal energies

give a quantitative measure of the certainty and an estimate of the quality of local registration accuracy.

In Sec. 3.3 the marginal probabilities that are obtained for the coarse image representation of supervoxels are used to estimate a displacement vector (or segmentation label) for every voxel \mathbf{x} with 3D coordinates $\mathbf{x} = \{x, y, z\}$. Our aim is to find the mode of the underlying distribution of potential displacements, $\mathcal{L}_p(\mathbf{x})$, which can be the full set of \mathcal{L} or a probable subset. The proposal solutions, \mathcal{L}_p , contain the information of all supervoxels v , which overlap at voxel \mathbf{x} . This step therefore fuses (or combines) the information of multiple complementary registration graphs. For the application of motion estimation, a continuous valued displacement vector is found by the non-parametric mean-shift algorithm (Comaniciu and Meer, 2002). Estimating the mode instead of the mean or median of the distribution \mathcal{L}_p makes our approach robust to outliers and multi-modal distributions of displacements. It also enables us to find a subvoxel optimum of the deformations (even though the space \mathcal{L} is confined to quantised displacements). When propagating segmentation labels, we do not search for the mode of the displacements, but instead their associated segmentation label. This can be seen as a label fusion step. However, instead of combining information from multiple atlases (as done e.g. in (Asman and Landman, 2013)), we fuse information from different potential transformations for a single atlas.

3.1. Supervoxel graph

Supervoxel clustering performs an over-segmentation of an image that respects image boundaries. Supervoxels remove the largely redundant intensity information of voxels within homogeneous areas, which are likely to belong to the same object. They provide a more compact image representation with little loss of detail.

For purposes of illustration, we adapt a recent algorithm, “simple linear iterative clustering” (SLIC) (Achanta et al., 2012) for supervoxel clustering. The algorithm is designed to create approximately equal-sized supervoxels $\mathbf{S} = \{S_1, S_2, \dots, S_K\}$. It starts from a set of equally spaced seed cluster centres with specified distance s (which determines the number of supervoxels $K \approx N/s^3$). The distance d_{ik} between a voxel $p_i = [l_i, x_i, y_i, z_i]^T$ (where l is the intensity value) and a cluster centre $S_k = [l_k, x_k, y_k, z_k]^T$ is given by the following equations:

$$\begin{aligned} d_{ik}^v &= |l_k - l_i| \\ d_{ik}^{xyz} &= \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2 + (z_k - z_i)^2} \\ d_{ik} &= d_{ik}^v + \frac{m}{s} d_{ik}^{xyz} \end{aligned} \quad (1)$$

The weighting m determines the compactness of the clusters, where higher values result in more regularly shaped supervoxels. It is assumed that the spatial extent of a supervoxel lies within a compact search region R of spatial extent $2s \times 2s \times 2s$. Therefore each voxel p_i has to be compared only to all centres, which are within R and is subsequently assigned to the closest cluster: $S(p_i) = \arg \min_{S_j \in R} d_{ij}$.

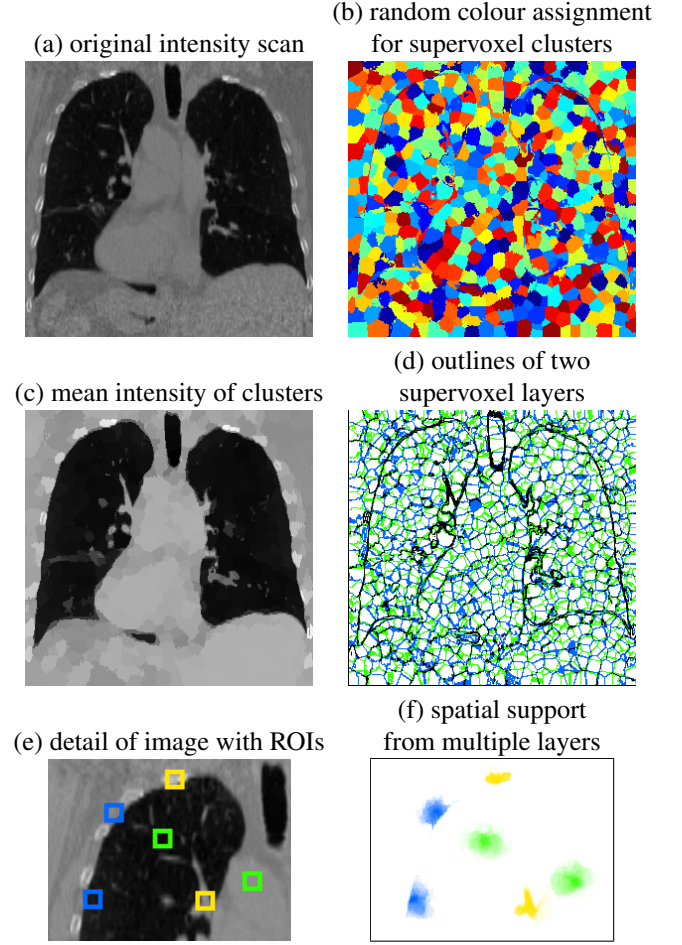


Figure 1: Example of multi-layer supervoxel representation. (a) Coronal slice of the original CT volume. (b) Supervoxel segmentation of (a) using a random colour assignment. (c) The abstracted image representation of (a) when assigning the mean intensity value to each cluster. This demonstrates the ability of supervoxels to preserve edges and small-scale structures. (d) The outlines of two supervoxel layers are shown together. Edges that are present in both layers are shown in black, indicating areas, e.g. the surface of the lungs, which are consistently clustered in both layers. (e) Regions of interest indicating small structures in yellow; homogenous regions in green; and locations close to boundaries in blue. (f) The superposition of supervoxels at these locations from 15 layers demonstrates the adaptive local support. The support region for small prominent structures (here a rib and a lung vessel) is tightly bound, for homogenous regions it is close to a Gaussian, and it adheres to image boundaries.

After one pass over all pixels, the cluster centres are recomputed. This process is repeated until the clusters no longer change. An example output of the method is shown in Fig. 1 (a-c). It may be seen that supervoxels can preserve image details and edges well.

3.1.1. Multiple layers of supervoxels

A single supervoxel clustering can only approximately represent the underlying image data. The clustering may be inconsistent in regions that are homogeneous or of gradually changing intensity. Furthermore, assigning the same displacement vector to every voxel in a supervoxel, is a limitation in the context of medical image registration, because deformations are in general not piece-wise constant. One of the main contributions of

our work (Heinrich et al., 2013b) is to use multiple layers of supervoxels (in contrast to e.g. (Lei and Yang, 2009)) to obtain a piecewise smooth motion model for accurate deformable registration. In Sec. 3.2.1, we present a way to optimise a registration cost function and obtain marginal distributions for each model separately. The combination of marginals resulting from different layers for the same voxel location is discussed in Sec. 3.3.

To create multiple layers of supervoxels, the SLIC algorithm is run several times with slightly different initialisations (random offset of seed locations with magnitude $\leq s/2$). Figure 1 (d-f) demonstrates this concept. Two different layers of supervoxels are shown in coronal view. The linear combination of the spatial layout of supervoxels of 15 layers for six regions of interest (homogenous, boundary, small feature) are shown in Fig. 1 (f). The close relation to the joint bilateral filter (Kopf et al., 2007), and the ability to model smooth regions and at the same time preserve edges can clearly be seen. The choice of parameters for the supervoxel creation are further discussed in Sec. 5.1.

Minimum spanning trees: In our graphical model the nodes, $v \in \mathcal{V}$, are represented by supervoxels and edges are found which connect neighbouring supervoxels and are therefore more likely to co-vary in their motion. Different neighbourhoods, \mathcal{N} , could be considered to connect all supervoxels in any layer, \mathcal{G} . Following our previous work (Heinrich et al., 2013a), we choose a minimum spanning tree (where edge costs are defined by the supervoxel distance d_{ik} as in Eq. 1). The motivation for this choice is that it provides a low computational complexity for the inference and enables us to find a globally optimal solution for each graph. The graph is built in the following way. First, a k-nearest neighbour graph is found by connecting every supervoxel to every other supervoxel, which shares at least one voxel in the immediate 26-neighbourhood. This guarantees us to find a spanning graph, which may, however, still contain loops. Second, using Prim’s algorithm (Prim, 1957) only edges that are necessary to form a spanning tree with lowest total edge cost are retained.

3.2. Registration Model

Deformable image registration using discrete optimisation can be formulated using a graph structure in which nodes $v \in \mathcal{V}$, correspond to control points of a transformation model (Glocker et al., 2008a). The space of potential displacements, \mathbf{d} , for each control point is discretised with a quantisation step of q yielding a three dimensional displacement space of $\mathbf{d} \in \mathcal{L} = \{0, \pm q, \pm 2q, \dots, \pm d_{\max}\}^3$. For a first order approximation, all voxels within the same supervoxel, denoted by $\eta(v)$, are assumed to follow the same constant motion vector. The cost, \mathcal{S} , for displacing a supervoxel, v_p , between target scan, I , and moving scan, J , can be evaluated independently for each node by adding the point-wise image similarity of all voxels in spatial proximity $\mathbf{x} \in \eta(v)$. Since no derivative of the similarity term is required, any point-wise metric can be used. In this work, we use the sum of absolute differences of thresholded intensity

gradients, $\nabla I'$, and thresholded intensities, I' (SADG):

$$\mathcal{S}(v_p, \mathbf{d}_p) = \sum_{\mathbf{x}_j \in \eta(v_i)} |I'(\mathbf{x}_j) - J'(\mathbf{x}_j + \mathbf{d}_j)| + |\nabla I'(\mathbf{x}_j) - \nabla J'(\mathbf{x}_j + \mathbf{d}_p)| \quad (2)$$

Prior to applying the thresholds, the image intensities are normalised between their first and 97th percentile. Similar to the normalised gradient fields introduced by (Haber and Modersitzki, 2007), SADG is less affected by locally changing image contrast than sums of squared differences (SSD).

Smooth transformations can be ensured by avoiding large differences in displacement vectors for neighbouring, connected nodes with a regularisation penalty, \mathcal{R} . In this work, absolute differences (known as total variation regularisation (Rudin et al., 1992)) 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:

$$\mathcal{R}(v_p, \mathbf{d}_p, \mathbf{d}_q) = \sum_{(p,q) \in \mathcal{N}} \frac{|\mathbf{d}_p - \mathbf{d}_q|}{\frac{s}{m}|l_p - l_q| + \|\mathbf{x}_p - \mathbf{x}_q\|} \quad (3)$$

A weighting parameter λ sets the influence of the regularisation in a combined energy function, $E(\mathbf{u})$, where $\mathbf{u}(\mathbf{x})$ describes the displacement field between two images. We aim to select the best displacements, \mathbf{d} , for all supervoxels in the image, by minimising:

$$E(\mathbf{u}) = \sum_{v \in \mathcal{G}} \mathcal{S}(v_p, \mathbf{d}_p) + \lambda \mathcal{R}(v_p, \mathbf{d}_p, \mathbf{d}_q) \quad (4)$$

3.2.1. MAP and marginal optimisation

Minimising the energy of Eq. 4 for arbitrary graph structures is NP-hard. For a tree graph, dynamic programming can be employed to find the global optimum of Eq. 4 (Felzenszwalb and Huttenlocher, 2005; Veksler, 2005) and thus the optimal displacement for each node. This concept has been applied successfully in our previous work on deformable lung registration (Heinrich et al., 2012, 2013a). Starting from the arbitrarily chosen root node, the full depth of the tree is explored by arranging all other nodes into ascending order of tree depth. At each node p , a message vector \mathbf{m}_p (outgoing message) containing the cost for the best displacement \mathbf{d}_p^* , given the displacement \mathbf{d}_q of its parent node q and the messages of all its children c (incoming messages), can be found by evaluating:

$$\mathbf{m}_p(\mathbf{d}_q) = \min_{\mathbf{d}_p} \left(\mathcal{S}(\mathbf{x}_p, \mathbf{d}_p) + \lambda \mathcal{R}(\mathbf{d}_p, \mathbf{d}_q) + \sum_c \mathbf{m}_c(\mathbf{d}_p) \right) \quad (5)$$

For any leaf node in the tree, Eq. 5 can be evaluated directly (since there are no incoming messages). Subsequently, the tree is traversed from its leaves to the root node (inward or forward pass). The message vector of (only) the root node then contains the exact marginal energies $\mathbf{m}^E(v_p, \mathbf{d}_p)$. The lower envelope technique of (Felzenszwalb and Huttenlocher, 2006) is used to evaluate Eq. 5 efficiently in $6|\mathcal{L}|$ calculations. The best displacements can be found by replacing min with argmin in Eq. 5 and then traversing the tree from the root down to leaves. This

Algorithm 1: Calculation of min-marginals using belief propagation.

```

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

```

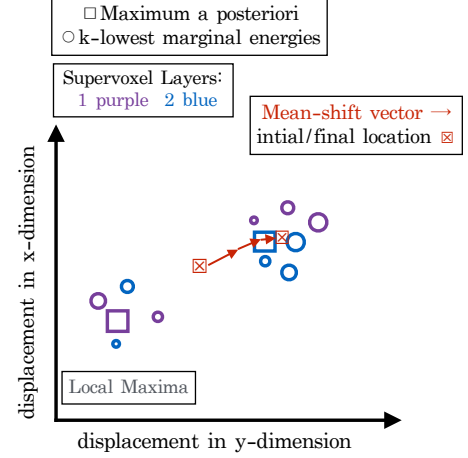


Figure 2: Mean-shift based motion estimation shown for two supervoxel layers and a two-dimensional displacement space. In this example, $k = 6$, $|\mathcal{G}| = 2$ and three iterations are shown. The size of the markers represents their probabilistic weight. Our proposed method is robust against local optima and finds the strongest optimum correctly with subvoxel accuracy.

is used in classical registration approaches, which only find the most probable transformation, viz. the maximum a posteriori (MAP) solution (Felzenszwalb and Zabih, 2011).

The marginal distribution for each node can be obtained by another pass of belief propagation (BP), but this time in the opposite direction. Algorithm 1 gives pseudo-code to obtain the min-marginal distributions. We introduce a second array, which stores the marginal distributions. During the backward pass, when evaluating a message send from the parent node to one of its children, the incoming messages from all other nodes have to be subtracted (see 3. in Algorithm 1). An advantage of using message passing, as compared to graph cut approaches is that marginal distributions for each node can be obtained directly. These can then be used to quantify the local uncertainty of the registration (for a proof, see (Kohli and Torr, 2008)).

The use of a spanning tree greatly reduces the computational complexity of the regularisation. However, some nodes that may have correlated displacements may not be connected. To overcome this limitation, we describe in the next section how several displacement proposals with quantified uncertainty, obtained from multiple comprehensive graph representations, can be combined to marginalise over these effects for improved segmentation propagation and motion estimation.

3.3. Mode seeking for proposal fusion

To date, few attempts have been made to directly employ the benefits of uncertainty estimates from discrete optimisation, and to the best of our knowledge min-marginals have not been used in medical image registration except in (Heinrich et al., 2013d) and (Parisot et al., 2014). 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 belief propagation (BP) (Tarlow and Adams, 2012). In our case, the use of BP on a tree enables the estimation of

a similarly good approximation of the marginals compared to loopy BP with greatly reduced memory demands and computation time.

Several challenges are commonly encountered in deformable registration (in particular for inter-patient registration), which can cause the occurrence of local uncertainties of the transformations. These include missing one-to-one correspondences of anatomies across subjects, mismatch of structures due to local minima of the registration, and untrustworthy image information due to acquisition noise and artefacts.

In the following, we show how mode seeking approaches can be used to compensate (in part) for inaccuracies by fusing many potential displacement proposals, which are obtained by independently optimising over each layer of supervoxels (as detailed in the previous section).

Mean-shift motion estimation: In the context of motion estimation, the proposals directly relate to the motion vector which is being estimated. Since real physical motion is not necessarily constrained to discrete voxel displacements, we would like to obtain an estimate of the underlying subvoxel mode of the proposal vectors. This can be achieved using the mean-shift mode seeking algorithm (Cheng, 1995; Comaniciu and Meer, 2002). An illustration of our approach is shown in Fig. 2.

For every voxel, we obtain a subset of more likely displacement vectors \mathcal{L}_p . These are the k displacements with lowest associated cost $\mathbf{m}^E(\mathbf{v}(\mathbf{x}), \mathbf{d})$ from each of the $|\mathcal{G}|$ layers. We initialise the algorithm with the mean motion vector $\mathbf{d}^0 = \bar{\mathbf{d}} = \frac{1}{|\mathcal{L}_p|} \sum_{i \in \mathcal{L}_p} \mathbf{d}_i$. For a number of iterations this vector is replaced by a weighted average over the distribution of all vectors using a Gaussian kernel function K . This yields a new estimate \mathbf{d}^{t+1} .

$$\mathbf{d}^{t+1} = \frac{\sum_{i \in \mathcal{L}_p} K(\mathbf{d}_i - \bar{\mathbf{d}}) \mathbf{d}_i}{\sum_{i \in \mathcal{L}_p} K(\mathbf{d}_i - \bar{\mathbf{d}})}, \text{ with } K(y) = \alpha \exp\left(-\frac{y^2}{2\sigma^2}\right) \quad (6)$$

Not all elements in \mathcal{L}_p have equal probability. They are weighted by $\alpha = \frac{k-r}{k}$, where $0 \leq r < k$ is the rank after sorting

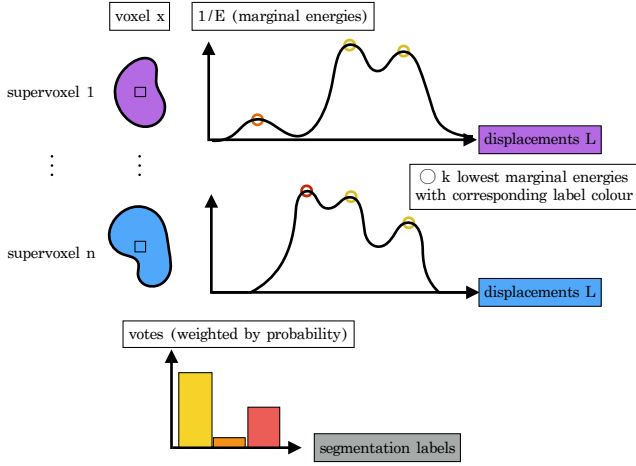


Figure 3: Visual explanation of our supervoxel displacement fusion. For every voxel \mathbf{x} , the $|\mathcal{G}|$ supervoxels, for which \mathbf{x} lies in $\eta(v)$ are considered. The marginal distribution over displacement energies for the supervoxels of interest is shown as a 1D vector. It is obtained for each layer of supervoxels by globally optimal belief propagation. The $k = 3$ displacements with lowest marginal energy (possibly several local optima) are selected. Using their corresponding segmentation labels (red, orange or yellow) from the atlas a voting, using normalised negated ranks according to Eq. 8, is performed for each voxel. Finally the segmentation label with maximal votes is selected (here yellow).

be weighted by their probability. The probability $p(\mathbf{x}, \mathbf{d})$ for each voxel $\mathbf{x} \in \eta(v_j)$ that is part of a supervoxel v_j and each displacement $\mathbf{d}_j \in \mathcal{L}$ can be directly obtained from the min-marginal energies $\mathbf{m}^E(v_j, \mathbf{d}_j)$ (see also (Kohli and Torr, 2008)):

$$p(v_j, \mathbf{d}_j) = \exp(-\beta \cdot \mathbf{m}^E(v_j, \mathbf{d}_j)) \quad (7)$$

We have used Eq. 7 in (Heinrich et al., 2013d) to find weightings for displacements (and their corresponding segmentation labels). However, the parameter β is difficult to determine empirically and is dependent on the variance of the underlying registration cost function. Instead, it is beneficial to consider only a subset of most probable vectors \mathcal{L}_p as in Eq. 6. The probability is then defined by its normalised negated rank:

$$p(v_j, \mathbf{d}) = \frac{2(k - r)}{k(k + 1)} \quad (8)$$

Where $r \in \{0, 1, \dots, k - 1\}$ is the rank obtained after sorting the k displacements with lowest energies $\mathbf{m}^E(v(\mathbf{x}), \mathbf{d})$ part of \mathcal{L} in an ascending order. The setting of k will be discussed in Sec. 5.1. 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. This procedure is visually presented in Fig. 3.

3.4. Symmetric registration

For deformable motion estimation, a symmetric, unbiased transformation and a one-to-one mapping of voxels across scans is desirable and physiologically expected. When estimating inter-subject deformations, symmetry and topology preservation may not always be realistic, however, such constraints can be useful for population studies. Hard constraints have been applied e.g. by (Sotiras and Paragios, 2012) and (Rueckert et al., 2006), but they limit the flexibility of each transformation and the deformation field has to be estimated as a composition of multiple small registration steps. Here, we follow previous work, e.g. (Muenzing et al., 2012b) or (Yushkevich et al., 2012) and perform a post-processing step, which ensures a one-to-one mapping. We use a similar concept to (Christensen and Johnson, 2001), which we have presented in (Heinrich et al., 2013c). Given the forward and backward displacement fields \mathbf{u}^n and \mathbf{v}^n respectively (which are independently calculated), we obtain two inverse consistent fields by iteratively updating the following equations:

$$\begin{aligned} \mathbf{u}^{n+1} &= 0.5(\mathbf{u}^n - \mathbf{v}^n(\mathbf{x} + \mathbf{u}^n)) \\ \mathbf{v}^{n+1} &= 0.5(\mathbf{v}^n - \mathbf{u}^n(\mathbf{x} + \mathbf{v}^n)) \end{aligned} \quad (9)$$

Empirically, we found that 10 iterations are sufficient to reduce the inverse consistency error (Christensen and Johnson, 2001) to insignificantly low values. In contrast to hard constraints, our approach has the advantage that it does not limit the maximum capture range of \mathcal{L} .

4. Materials and Experiments

We have evaluated our registration framework for two challenging medical image registration tasks.

4.1. Description of datasets for experimental validation

First, we estimate intra-patient motion during the respiratory cycle for the DIR-lab 4DCT dataset (Castillo et al., 2009). The dataset has been made publicly available by the Anderson Cancer Center, University of Texas. It consists of ten thoracic computed tomography (CT) scans of cancer patients undergoing 4DCT scanning for radiotherapy treatment planning. The time-points at maximal inhale and exhale (T00 and T50) are non-rigidly registered to estimate the motion of the tumour and all other internal lung structures in order to more accurately determine sufficient radiation margins. In cases where the tumour might be close or attached to the lung pleura, it might furthermore be important to estimate the sliding motion, commonly present during normal breathing, and therefore also calculate the motion of the thoracic cage (including the ribs). The resolution of these scans varies from 0.97 to 1.16 mm in-plane resolution, the slice thickness is 2.5 mm. Scans #6-10 were cropped to exclude regions outside the body, which resulted in volumes of roughly 7 million voxels. 300 manually defined landmark pairs have been provided by (Castillo et al., 2009) for evaluation of image registration algorithms, with intra-observer accuracy of around 1 mm on average.

Second, T1-weighted magnetic resonance images (MRI) of the brain from forty subjects of the LPBA40 (Shattuck et al., 2008) dataset are used for atlas-based segmentation propagation. The scans are manually parcellated into 56 anatomical regions by trained raters. The images have an isotropic spatial resolution of 1 mm. The dimensions of all images are $181 \times 217 \times 181$ voxels. The pre-processing of the data follows (Klein et al., 2009), who performed an extensive evaluation of 14 non-linear registration algorithms for the same dataset. To be comparable to this study, we only perform pair-wise registrations and label propagation.

4.2. Description of variants of our registration algorithm

We report a number of experiments to study the influence of our presented methodological contributions. In total six variants of our algorithm are used:

- (A) We run the registration without inference of regularisation of the motion field and apply the mean-shift mode seeking approach to the candidate motion vectors from the $|\mathcal{G}|$ different layers. This approach is similar to previous work on supervoxel matching, e.g. (Wang and Yushkevich, 2013).
- (B) We add the displacement field regularisation using the forward pass of dynamic programming. The arithmetic mean of the $|\mathcal{G}|$ maximum a posteriori (MAP) solutions is used as final motion vector. This was the same strategy, we employed in our previous work (Heinrich et al., 2013b).
- (C) A second pass of belief propagation is performed to obtain the marginal distribution over \mathcal{L} for every supervoxel and every layer. These proposals are used in the mean-shift mode seeking approach to obtain a voxel-wise displacement field. For the LPBA40 data, the segmentation label at the location in the manually annotated scan, pointed at by this vector is assigned.

- (D) A symmetric solution is obtained by first estimating forward and backward transform separately (using marginal distributions and mean-shift) and finding the displacement field that minimises their discrepancy (see Sec. 3.4).
- (E) The same as (D) but using loopy BP (as described in (Felzenszwalb and Huttenlocher, 2006) without multi-scale grids) for the inference of regularisation on a much more densely connected graph with loops, and with symmetry.
- (F) The marginal distributions of all layers (over \mathcal{L}) are directly used to obtain the weights for a voxel-wise label fusion (by a weighted majority vote, see Fig. 3). Note, this variant only applies to segmentation propagation and uses the same inference as in (C) based on MSTs.

The sensitivity of the setting of the most important parameters will be studied in Sec. 5.1.

5. Results and Discussion

In the following, we present and discuss the results achieved with our algorithm (and its variants) for the two datasets. Since both registration tasks have a standardised evaluation, which has been used in previous published work, we can compare the presented displacement fusion strategy to the state-of-the-art. The experiments on 4D-CT lung motion estimation are evaluated in terms of target registration error (TRE) for 300 manually selected landmark pairs for every scan. For the second tasks of propagating brain segmentations across MRI scans of different subjects, we evaluate the Jaccard coefficient (or union overlap). This is defined as $J = (A \cap M) / (A \cup M)$ (calculated separately for each of the 56 label regions) between automatic and manual segmentations A and M . Additionally, the smoothness of the obtained transformations is analysed visually and quantitatively. The standard deviation of the determinant of the Jacobian matrix (of the final displacement field \mathbf{u}) is used as a measure of the complexity of the transformations. Examples of the displacement fields are shown to confirm the overall plausibility and demonstrate in detail the ability of our approach to deal with spatially varying deformation smoothness.

5.1. Parameter choice

Five main parameters have to be set in our method, which will be made publicly available at <http://www.mpheinrich.de>. $|\mathcal{G}|$, the number of overlapping supervoxel layers; λ , the weighting of the total variation regularisation; s , the step-size, which determines the average size of the supervoxels; k , the number of displacements with lowest energy to be considered for every supervoxel; and the definition of the full displacement space \mathcal{L} . We study the influence of these parameters for one case of the DIR-lab and three registrations of the LPBA40 dataset.

We found the setting of the compactness parameter (m in Eq. 1) to be fairly insensitive. Best results are obtained for $\mathbf{m} = \mathbf{1}$, but the TRE increases by less than 0.08 mm for $0.25 \leq m \leq 8$. The displacement space \mathcal{L} is quantised with a step-size of 2 voxels and a maximum capture range (for the lung data) of

$d_{\max} = 14$ voxels has been chosen, which results in $|\mathcal{L}| = (2 \times 7 + 1)^3 = 3375$ displacement labels. The maximum capture range for the LPBA40 data is chosen to be 10 mm (with a quantisation of 2 voxels). For all experiments the intensity and intensity gradient thresholds for the SADG similarity measure (see Eq. 2) are set to 0.25 and 0.02 respectively.

The influence of the remaining four free parameters is studied in more detail in the following. When changing the other parameters, the regularisation weight was set to $\lambda = 0.4$, the number of best displacement per supervoxel to $k = 9$ and the number of layers to $|\mathcal{G}| = 5$. We used 4 mean-shift iterations in all experiments.

Figure 4 (a) shows the accuracy of segmentation propagation (using variant F) and motion estimation (using variant D) in terms of Jaccard overlap and TRE for our method in dependency of the number of supervoxels. It can be seen that at least $|\mathcal{G}| = 5$ layers of supervoxels are necessary to have sufficient spatial overlap and to obtain a good voxel-wise approximation of motion vectors and segmentation labels. An increasing number of supervoxel layers leads in general to improved accuracy (at the cost of higher memory and computation time requirements).

The weighting parameter for the smoothness, λ , is varied over a large range of values (doubling each subsequent value) in Fig. 4 (b). The choice of λ is relatively insensitive, values from an interval of $0.1 \leq \lambda \leq 0.4$ can be used for both tasks. Since smoother transformations have a lower complexity and are therefore usually favourable, we select $\lambda = 0.4$.

The variation of the size k of the subset, \mathcal{L}_p , of most probable displacements for each layer can be seen in Fig. 4 (c). For motion estimation an optimum is found for $k = 9$ (i.e. in total $9 \times 5 = 45$ candidates are evaluated for each voxel), while the segmentation performance increases further for values of $k > 15$. This indicates that the ability of the mean-shift algorithm to accurately find the mode of the distribution is reduced (leading to slightly inferior registration accuracy), if the number of local optima is too large. However, for the weighted majority voting the choice of k is less important.

For the lung CT data, a supervoxel step-size of $s \approx 9$ mm is chosen (this translates into 9 voxels in plane and 4 voxels in the z-dimension) to provide a reasonable trade-off between spatial context and local flexibility (see Fig. 4 (d)). To be able to deal with the higher variability of small anatomical structures in the brain across subjects (compared to the lung experiment), we use a smaller supervoxel size of $s = 6$ mm.

5.2. Evaluation of CT lung registration

Estimation of lung motion is a challenging image registration problem. The absolute motion is in general quite large (>25 mm) and there is a large discrepancy of the magnitude and direction of the motion inside and outside of the lungs. This sliding motion, which has been studied in many previous articles, e.g. by (Schmidt-Richberg et al., 2012b), would require a locally varying smoothness assumption with a discontinuity at the interface between lung surface and rib cage. Currently, most approaches that achieve the highest accuracy

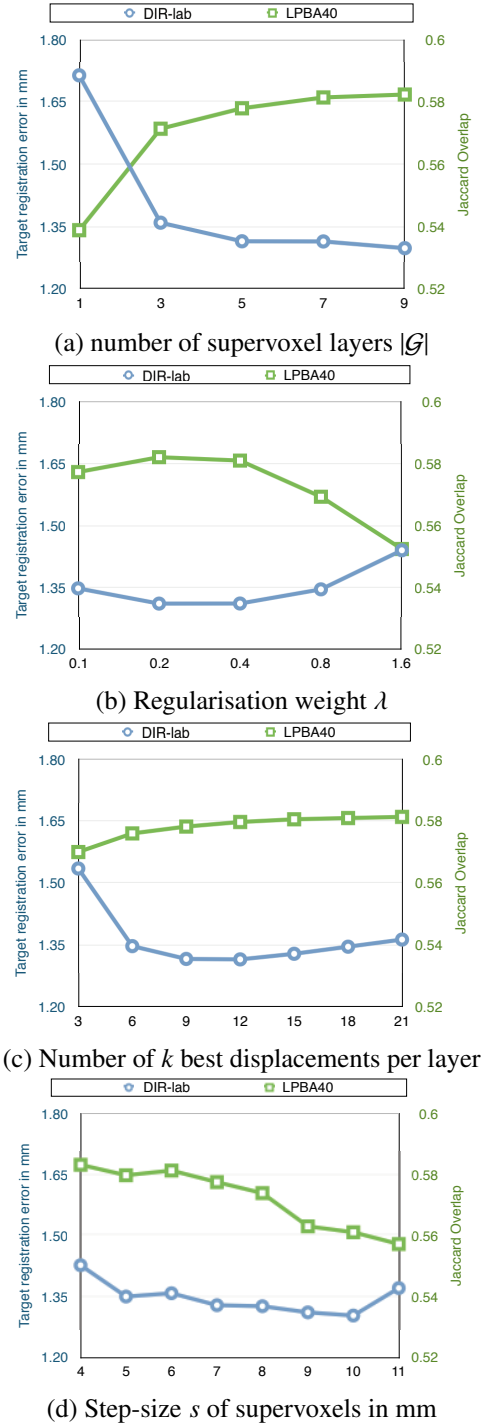


Figure 4: Parameter variation to find the optimal settings for our method. One case (#7) of the DIR-lab lung CT data and three registrations for the LPBA40 scans are studied. (a) At least 5 layers of supervoxels should be used. More complementary layers improve the results further at the cost of higher computational requirements. (b) A good performance (high Jaccard overlap and low TRE) is achieved for a weighting of the regularisation between $\lambda = 0.1$ and $\lambda = 0.4$. (c) A number of $k = 9$ of displacements with lowest marginal energy per supervoxel (considered for mode seeking) yields the lowest TRE. The segmentation overlap levels off for values larger than $k = 12$. (d) The average distance between two supervoxel centres yields good results at $s = 9$ mm for the lung and $s = 6$ mm for the brain data.

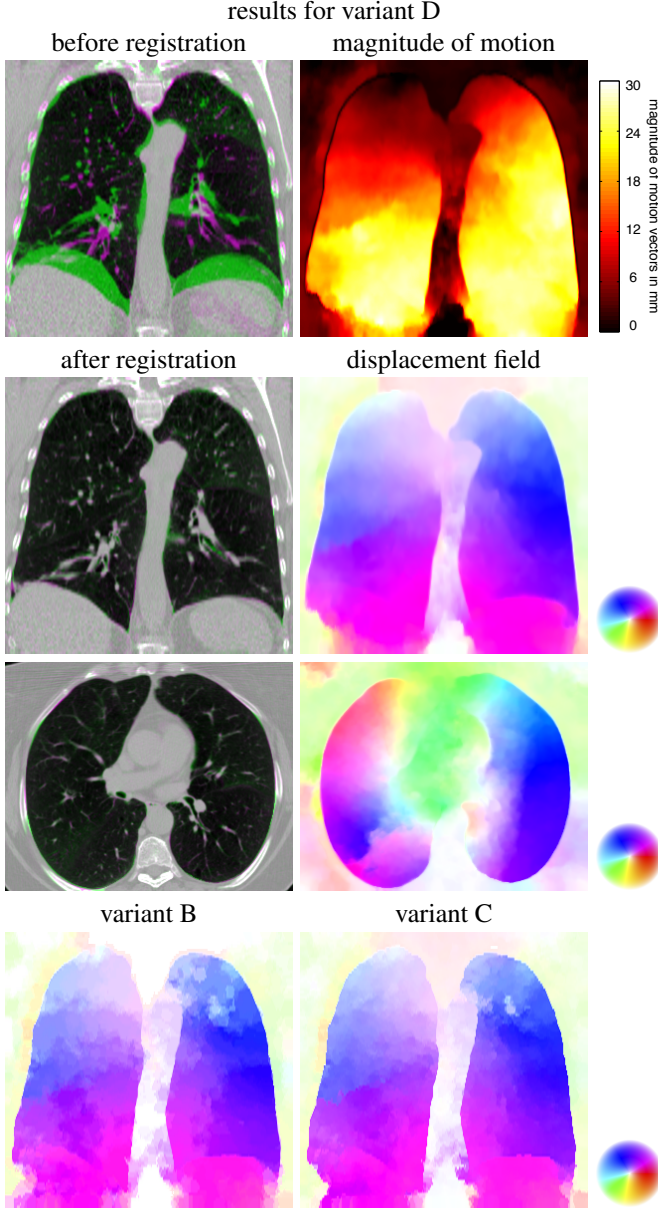


Figure 5: Registration result for case 8 of 4D CT dataset shown for our symmetric registration approach (D). The first column shows an overlay of inhale (green) and exhale (magenta) phase before (in coronal view) and after registration in both coronal and axial view. The second column displays the motion magnitude in mm, as well as the colour-coded displacement field (using an HSV-colour representation for vector orientation) for coronal and axial planes. The displacement fields demonstrate that the sliding of the lungs is well preserved. Comparing variants B and C in the last row, the smoothness of the motion increases by each including marginal estimates and the symmetric constraint. Furthermore, occasional outliers are reduced.

algorithm	B	C	D	E
	no marginals	meanshift	symmetric	loopy BP
accuracy (TRE)	1.619 ± 1.13 mm	1.326 ± 0.87 mm	1.236 ± 0.82 mm	1.257 ± 0.87 mm
harmonic Energy	0.395	0.333	0.157	0.134
mean(Jacobian)	0.887	0.889	0.880	0.881
std(Jacobian)	0.292	0.251	0.126	0.126
fraction of neg. J.	5.9E-03	4.0E-03	2.0E-05	2.9E-05

Table 1: Comparison of metrics for smoothness of transformations for lung registrations (**DIR-lab**). All values have been computed inside a masked region of the lungs (to eliminate differences due to cropping). It can be seen that for employing marginals using the meanshift mode finding reduces the harmonic energy. A further substantial reduction in complexity is achieved when using a symmetric variant. Using loopy belief propagation (LBP) (with a more densely connected graph) and a symmetric constraint may further reduce harmonic energy, however it comes with much higher computational demand. The results compare well to the ones published in the literature by (Werner et al., 2014), who reported standard deviation of Jacobians of 0.131 and a TRE of 1.24 ± 0.91 mm. More detailed evaluations can be found in the Appendix in Table 3.

algorithm	C	D	E	F
	meanshift	symmetric	LBP sym.	voting
accuracy (Jaccard)	0.558 ± 0.07	0.555 ± 0.07	0.558 ± 0.07	0.580 ± 0.07
harmonic Energy	1.862	0.345	0.297	
mean(Jacobian)	1.025	1.013	1.009	
std(Jacobian)	1.241	0.434	0.392	
fraction of neg. J.	8.7E-02	1.8E-03	9.1E-04	

Table 2: Comparison of metrics for smoothness of transformations for brain registration (**LPBA40**), computed inside a masked region of the brains. The best accuracy $J = 0.580$ is obtained for variant F, using marginal voting. Nearly the same overlap of ≈ 0.56 is achieved when employing the meanshift variants (C,D,E) with increased smoothness from C to D and D to E. The results for using LBP **without symmetry** are slightly more accurate $J = 0.563 \pm 0.07$ but less smooth than variants D or E with a harmonic energy of 0.932 , a standard deviation of Jacobians of 0.821 , and a fraction of negative Jacobians of 4.1×10^{-2} .

for the DIR-lab dataset (Hermann and Werner, 2013; König and Rühaak, 2014), deal with this problem by segmenting the lungs and masking out the intensities of rib-cage and other body parts. Yet, estimating the motion over the whole domain is clinically useful for radiotherapy of lung cancer (where the tumour might be attached to the pleura) and the assessment of breathing disorders. Most approaches to this problem have used a manual or automatically detected segmentation of the thoracic cage (Schmidt-Richberg et al., 2012b; Risser et al., 2012; Vandemeulebroucke et al., 2012). Our approach relies solely on the over-segmentation automatically available from the supervoxel clustering, and the mean-shift mode finding, which is robust (up to a certain degree) to multi-modal distributions of displacement vectors (this is important at locations where a supervoxel crosses a motion boundary). It can be clearly seen from the visual example in Fig. 5 that our method is able to preserve motion boundaries without requiring any manual interaction or prior anatomical knowledge. The method would therefore be very suitable to quantify sliding motion e.g. following (Amelon et al., 2013).

A quantitative evaluation of our results over all ten cases is

shown in Table 1 and Table 3. The average run-time of our algorithm is less than 2 minutes using a single CPU core (40 sec. for similarity computation and 20 sec. each for SLIC, inference and meanshift). We compare the registration error after applying the first five variants (A-E) of our method as discussed in Sec. 4.2. The settings used for these registration tasks are (as discussed above): $|\mathcal{G}|=5$, $\lambda=0.4$, $k=9$ (all approaches use the same supervoxel layers). It is obvious that omitting the regularisation (variant A), $\lambda = 0$, yields inaccurate registration results for most cases resulting in an average TRE of 2.82 ± 0.92 mm. Inferring regularity of the displacement field using dynamic programming and selecting the average MAP solution for the displacements over all layers reduces the TRE to 1.62 ± 1.13 mm. Note that this approach is similar to our previous work (Heinrich et al., 2013b) (apart from using another similarity metric and a slightly different displacement space), which achieved a TRE of 1.94 mm (evaluated for cases #6-10 only). Our proposed method, which uses the uncertainty estimates after two passes of belief propagation, yields a significant improvement over the MAP approach with an TRE of 1.33 ± 0.87 mm (a reduction of 0.3 mm). Statistical significance was tested using one-tailed Wilcoxon rank sum test for each scan pair individually, resulting in $p < 0.05$ for all cases. This result can be further improved to **1.24 ± 0.82 mm** when using the symmetric constraint (and estimating both forward and backward transformations). The obtained deformations (see Fig. 5 for example) are smooth for all areas besides the lung-rib interface where sliding motion is expected and correctly identified. The complexity of the transformations was calculated using the determinant of the Jacobian and the harmonic energy, see Table 1. The harmonic energy for variant D is on average 0.157 and ranges between 0.063 (for case #1) and 0.2690 (for case #8). The asymmetric approach yields approximately twice these values. When using a denser graph (with ≈ 14 edges per node) and loopy BP (variant E) the accuracy cannot be further improved, the complexity measured by Jacobians is also similar, and the harmonic energy only decreases slightly. Note that we found 5 iterations of LBP are sufficient and the regularisation weighting has been adapted to reflect the greater number of edges. Given that this inference requires a computation time of more than 10 minutes (compared to 10 seconds), it demonstrates that our tree approximation has good practical value.

Comparing our approach to state-of-the-art methods, we find that our results are superior in terms of TRE compared to approaches that estimate the full deformation fields for this dataset: 2.34 mm (Papiez et al., 2013), 1.55 mm (Schmidt-Richberg et al., 2012a), 1.43 mm (Heinrich et al., 2013a) and 1.41 mm (Hermann and Werner, 2013) (without masks). They do, however, not entirely match the performance previously achieved for masked registration: TRE of 0.99 mm and 0.95 mm (Hermann and Werner, 2013; König and Rühaak, 2014). However, as stated above, the estimation of the full motion (including the sliding behaviour) is important for clinical applications.

5.3. Evaluation of segmentation propagation for brain MRI

The LPBA40 dataset has been used in a comprehensive comparison study of 14 non-linear registration methods in (Klein et al., 2009), which enables us to directly compare our approach to 14 other algorithms for a total of 1580 one-to-one registrations. Note that in all experiments only pair-wise segmentation propagation is performed and no multi-atlas fusion has been used. An example registration/segmentation outcome is presented in Fig. 6. Key variants of our algorithm, as described in Sec. 4.2 are compared. The accuracy of variant B, which only estimates the most probable of all transformations, in terms of Jaccard overlap is 0.522 ± 0.074 . Because the use of dynamic programming only determines the overall cost for each layer, this solution essentially selects the best image representation model globally. Variant B would be placed in 7th position compared to all 14 algorithms. When calculating the full marginal distribution over all displacements and estimate the mode over them using the mean-shift algorithm (variant C), the overlap increases to $J=0.558 \pm 0.073$. Note that this approach estimates a displacement field and is equivalent to our motion estimation of the previous section. The obtained transformation is afterwards applied to the manually segmented volume to transfer the labels. A symmetric variant D achieves similar accuracy $J=0.555 \pm 0.072$ with much smoother deformations (see Table 2). When using loopy-BP for inference (variant E), the obtained results are slightly improved in terms of smoothness (difference significant $p < 0.01$) and accuracy (difference insignificant). Finally, we perform the same optimisation as for variant C, but this time remove the mean-shift motion estimation step and directly apply the uncertainty-based weights in a voxel-wise label fusion. The parameter k is set to 20 for this experiment (based on Fig. 4 c) where it was found that $k = \{12, \dots, 21\}$ gave similar accuracy). This method (variant F), which is useful in practice when the labelling of anatomies is the goal of registration, performs a probabilistic single-atlas label fusion and achieves the best results with an overlap $J=0.580 \pm 0.072$. Our approach compares favourably to algorithms tested by (Klein et al., 2009), see Fig. 7. The run time of our algorithm is similar to the experiments in Sec. 5.2, with about 90 seconds on a single CPU core, which is at least an order of magnitude faster than the other top performing algorithms. A further possible variant would be to use the label fusion approach (as in variant F) for the single MAP solution (variant B) independently obtained for each supervoxel layer. This substantially improves the overlap to $J=0.560$ close to the results of SyN, ART and variant C, but is still about 0.02 inferior to variant F, which employs uncertainty estimates. (Ou et al., 2014) have subsequently published results for the popular DRAMMS algorithm, which achieves $J=0.566$. We have recently published a method (Heinrich et al., 2014) using local cost aggregation together with a global smoothing that yields an overlap of $J=0.564$.

5.4. Registration error estimation using marginal distribution

Another valuable application of our method is the use of the uncertainty estimates to detect registration errors. After applying the adaptive mean-shift algorithm to the set of displacements \mathcal{L}_p for every voxel, we obtain an estimate of the

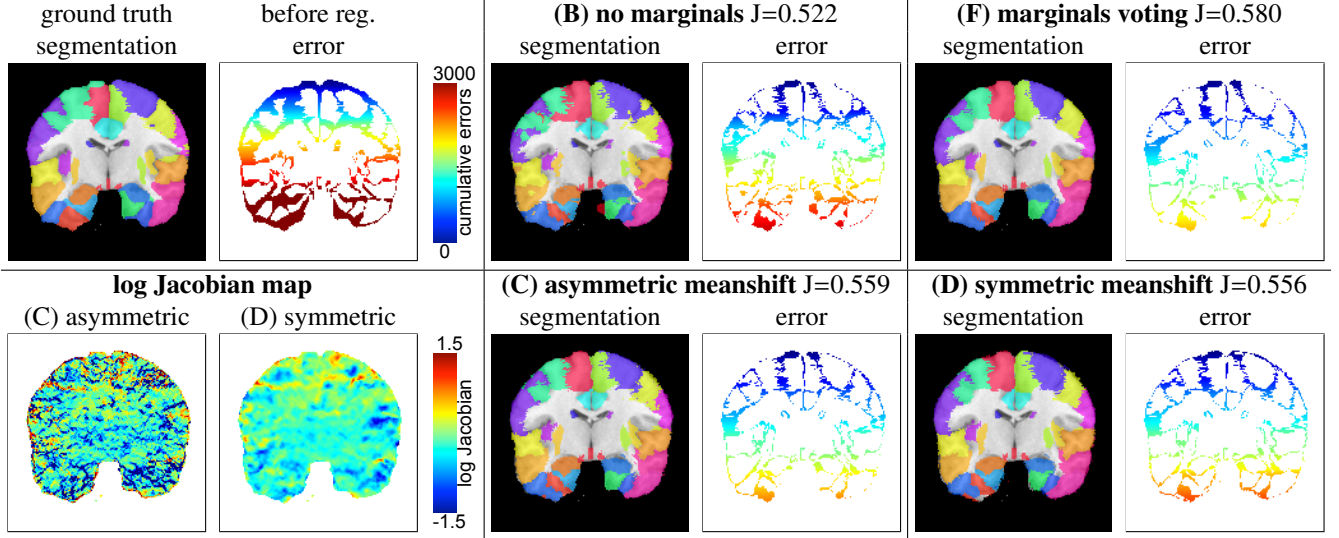


Figure 6: Visual examples for LPBA40 dataset for registration and segmentation propagation, between subjects 2 and 33, which results in median performance. The segmentations are noisier when not using uncertainty estimates (B) and some structures, e.g. left precentral gyrus ■ and left sup. frontal gyrus ■ are overestimated. Variants C and F, achieve lower errors next to boundaries between labels or at the brain surface (see e.g. the left inferior temporal gyrus ■). The cumulative sum of labelling errors within this coronal slice are shown in false colour. The complexity, shown using log Jacobian maps is low in homogenous regions of the white matter and higher in certain complex regions near cortical folds. The symmetric variant D yields much smoother transformations than C. Results for the best and worst registrations are shown in the supplementary material.

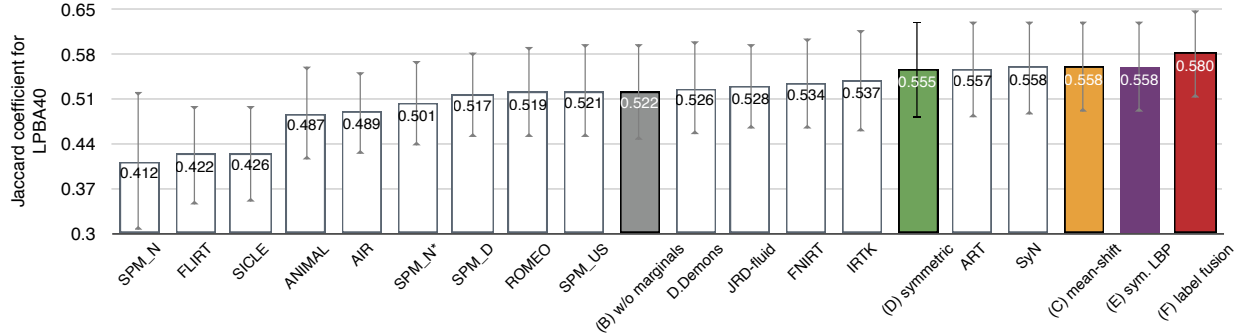


Figure 7: Volume overlap (Jaccard) averaged over 1560 registrations and 56 anatomical labels. The mean-shift motion estimation using uncertainty estimates from the presented inference (C) and its symmetric variant (D) achieve equivalently good results as the best performing methods (SyN and ART). Our proposed method employing multiple supervoxel layers and uncertainty estimates for voxel-wise label fusion (F) outperforms all previously tested methods from (Klein et al., 2009). Note, that the Jaccard score is calculated for each label individually and then averaged. These results are as published by Klein at: http://www.mindboggle.info/papers/evaluation_NeuroImage2009/tables/table_ROIxMethod_U0_LPBA40.html and not according to Eq. 4 in his paper. When only the MAP solution is considered (B) without taking uncertainties into account, the achieved accuracy would be ranked 7th in the comparison of 14 algorithms.

local standard deviation $\sigma(\mathbf{x})$ of the displacement distribution.⁸⁰⁸
To illustrate the potential utility of this, we performed the fol-⁸⁰⁹
lowing test for CT lung registrations. A smaller capture range⁸¹⁰
of $d_{\max} = 6$ voxels is used for the displacements and the lo-⁸¹¹
cal uncertainty of the registration of all inhale-exhale scans⁸¹²
is calculated. The displacement range would be sufficient for⁸¹³
scans with smaller motion (cases #1-5), but not for the exam-⁸¹⁴
ple shown in Fig. 8 (case #8). We split the patients into two⁸¹⁵
sets (odd and even number). The landmark error from the train-⁸¹⁶
ing set (even) is used to train a linear classifier, which separates⁸¹⁷
voxels into well or poorly aligned ($TRE > 2.5$ mm). We obtain a⁸¹⁸
value of $\sigma^2 = 2.13$ from plotting the thresholds of the ROC and⁸¹⁹
picking the location of maximal curvature (see example in Fig.⁸²⁰
8 at the bottom). Using this threshold enables us to differentiate⁸²¹
locations in the second testing set (odd cases), based on their

registration uncertainty, between good correspondences (mean
landmark TRE of 1.35mm) and poor correspondences (mean
TRE of 9.63mm). From the visual example in Fig. 8 int the top
row it can be seen that our method recognises local registra-
tion errors (high intensity in uncertainty map in the upper lobe)
even though they are not directly detectable based on image
similarity alone. The regression plot (green circles) shows that
 σ (which is measured in voxels) can be directly correlated to
the registration error TRE (in mm), which were obtained with
manual landmarks. It is therefore possible to detect errors in
the absence of a gold standard. When the displacement range
is increased to 14 voxels (as in our previous experiments) the
ground truth error and the uncertainty estimate are reduced to
very small values (blue squares).

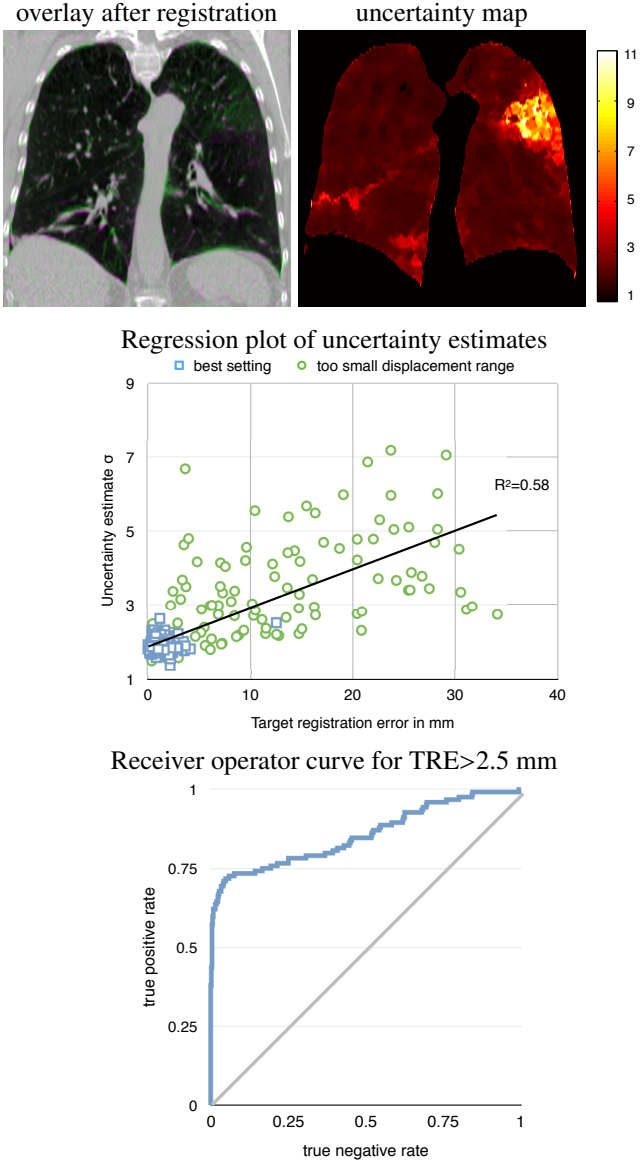


Figure 8: Registration uncertainty estimated by the local standard deviation $\sigma(\mathbf{x})$ of the adaptive mean-shift algorithm. Two scenarios are compared in this example for case #8 (representative coronal slices of 3D registration are shown): First, a registration with too small capture range $d_{\max} = 8$ voxels, for which the uncertainty map is shown on the left. Second, motion estimation with full range of displacements $d_{\max} = 14$ voxels. The xy-plot of $TRE(\mathbf{x})$ and $\sigma(\mathbf{x})$ at locations of manual landmarks indicates a correlation with R^2 value of 0.58 over both tests. Using the uncertainty estimates, $\sigma(\mathbf{x})$, for binary classification into aligned ($TRE < 2.5$ mm) and non-aligned voxels yields a good receiver operator curve with an area under the curve of 0.855. Here, the uncertainty estimates can be used to detect the registration errors in the upper right lobe, which are not easy to detect based on image similarity alone.

6. Conclusion

We have presented a new method for motion estimation and segmentation propagation, which together make two contributions to the field of medical image registration. First, we have proposed to represent the image domain by **multiple complementary representations** using layers of overlapping supervoxels. This overcomes some limitations of sparse image models by first estimating multiple global solutions to the registration problems and then choosing the locally best model. The inference of deformation smoothness on a sparse graph can be performed very efficiently using belief propagation with guaranteed convergence after two passes. Second, we use a large space of potential displacement vectors and calculate the marginal distribution of them for every node in the simplified graph. Instead of only considering the most probable (the maximum a posteriori, MAP) transformation, this enables us to take into account the spatial uncertainty of the registration.

To obtain voxel-wise approximations of the optimal motion vector or segmentation label from our sparse graphical model, we consider the most likely candidates from the displacement space and all complementary image representations. The mode of these candidates is found using the well-known mean-shift algorithm to estimate motion vectors; and a simply voting scheme for segmentation labels. Both fusion strategies take into account the uncertainty over the set of proposals.

The concepts we have presented have a beneficial impact to overcome a number of present challenges in medical image registration. The use of multiple layer of supervoxels enables us to represent the complex nature of deformations across medical scans with spatially varying smoothness. The suitability of our method to preserve discontinuous sliding motion at the lung surface, which is common during respiration, and concurrently estimate smooth deformations in other areas has been shown in our experiments. Prospectively, this could also be applied for cardiac motion estimation as an alternative to sparse free-form deformations (Shi et al., 2013). For segmentation propagation, the multi-layer approach improves the labelling accuracy for difficult anatomical structures and extends the popular concept of multi-atlas fusion to one-to-one image registration. Future work could further improve on this idea by using more powerful fusion strategies, e.g. non-local STAPLE (Asman and Landman, 2013), instead of our simple weighted voting.

Making use of marginal distributions over a large space of displacements has been shown to significantly reduce the target registration error for respiratory motion estimation. The mean-shift algorithm is able to accurately find a subvoxel motion vector based on the quantised displacements and is therefore able to mitigate one of the prevalent limitations of discrete registration. Other robust estimators, which reject outliers, could be studied for this purpose in future work. The obtained motion fields have a good overall regularity (especially when including a symmetry constraint), because the overlapping supervoxels provide a smooth transition between sets of displacement proposals. However, further improvements could be made by inferring regularity of deformations on a denser voxel graph using the concept of outer-planar graphs by (Batra et al., 2010) or

by considering only the small subset of probable displacements from the coarse scale. Similar ideas have been explored for stereo (Yang et al., 2010), optical flow (Glocker et al., 2008b) estimation and 3D medical image registration (Heinrich et al., 2011).

Using more connections per node than in the simplified tree model tends to improve the smoothness of transformations at the cost of higher complexity. In future work, the use of sequential tree-reweighted message passing (TRW-S) (Kolmogorov, 2006) could be of great interest. However, an open question for our model, which is not on a grid, would be a suitable automatic selection of monotonic chains that are required by TRW-S.

A further useful feature of our presented method is the fact that no (iterative) image warping scheme is necessary. Since only the original (non-transformed) images are used to calculate the similarity measure, approaches that rely on time-consuming feature extraction, e.g. (Ou et al., 2011), could be significantly speeded up.

Acknowledgments

The authors would like to thank EPSRC and Cancer Research UK for funding parts of this work within the Oxford Cancer Imaging Centre. I. Simpson was supported by the NIHR Queen’s Square Dementia BRU. We are thankful to the anonymous reviewers for their valuable comments and suggestions that greatly helped to improve this paper.

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., nov. 2012. SLIC superpixels compared to state-of-the-art superpixel methods. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 34 (11), 2274–2282.

Amelon, R. E., Cao, K., Reinhardt, J. M., Christensen, G. E., Raghavan, M. L., 2013. A measure for characterizing sliding on lung boundaries. *Annals of biomedical engineering*, 1–9.

Ashburner, J., Ridgway, G. R., 2012. Symmetric diffeomorphic modeling of longitudinal structural MRI. *Frontiers in neuroscience* 6.

Asman, A. J., Landman, B. A., 2013. Non-local statistical label fusion for multi-atlas segmentation. *Medical image analysis* 17 (2), 194–208.

Batra, D., Gallagher, A. C., Parikh, D., Chen, T., 2010. Beyond trees: MRF inference via outer-planar decomposition. In: *Computer Vision and Pattern Recognition (CVPR)*, 2010 IEEE Conference on. IEEE, pp. 2496–2503.

Cardoso, J. M., Leung, K., Modat, M., Keihaninejad, S., Cash, D., Barnes, J., Fox, N. C., Ourselin, S., 2013. Steps: Similarity and truth estimation for propagated segmentations and its application to hippocampal segmentation and brain parcellation. *Medical image analysis* 17 (6), 671–684.

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 (7), 1849.

Cheng, Y., 1995. Mean shift, mode seeking, and clustering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 17 (8), 790–799.

Christensen, G., Johnson, H., 2001. Consistent image registration. *IEEE Transactions on Medical Imaging* 20 (7), 568–582.

Cobzas, D., Sen, A., 2011. Random walks for deformable image registration. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2011*. Springer, pp. 557–565.

Comaniciu, D., Meer, P., 2002. Mean shift: A robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24 (5), 603–619.

Comaniciu, D., Ramesh, V., Meer, P., 2001. The variable bandwidth mean shift and data-driven scale selection. In: *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. Vol. 1. IEEE, pp. 438–445.

Coupé, P., Manjón, J., Fonov, V., Pruessner, J., Robles, M., Collins, D., 2011. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *Neuroimage* 54 (2), 940–954.

Datteri, R. D., Dawant, B. M., 2012. Automatic detection of the magnitude and spatial location of error in non-rigid registration. In: *Biomedical Image Registration*. Springer, pp. 21–30.

Felzenszwalb, P., Huttenlocher, D., 2006. Efficient belief propagation for early vision. *International Journal of Computer Vision* 70, 41–54.

Felzenszwalb, P. F., Huttenlocher, D. P., 2005. Pictorial structures for object recognition. *International Journal of Computer Vision* 61, 55–79.

Felzenszwalb, P. F., Zabih, R., 2011. Dynamic programming and graph algorithms in computer vision. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 33 (4), 721–740.

Ferrant, M., Warfield, S. K., Nabavi, A., Jolesz, F. A., Kikinis, R., 2000. Registration of 3D intraoperative MR images of the brain using a finite element biomechanical model. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2000*. Springer, pp. 19–28.

Forsberg, D., Rathi, Y., Bouix, S., Wassermann, D., Knutsson, H., Westin, C.-F., 2011. Improving registration using multi-channel diffeomorphic demons combined with certainty maps. In: *Multimodal Brain Image Analysis*. Springer, pp. 19–26.

Glocker, B., Heibel, T. H., Navab, N., Kohli, P., Rother, C., 2010. Triangleflow: Optical flow with triangulation-based higher-order likelihoods. In: *Computer Vision—ECCV 2010*. Springer, pp. 272–285.

Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N., 2008a. Dense image registration through MRFs and efficient linear programming. *Medical Image Analysis* 12 (6), 731–741.

Glocker, B., Paragios, N., Komodakis, N., Tziritas, G., Navab, N., 2008b. Optical flow estimation with uncertainties through dynamic MRFs. In: *Computer Vision and Pattern Recognition, CVPR 2008*. IEEE, pp. 1–8.

Glocker, B., Sotiras, A., Komodakis, N., Paragios, N., 2011. Deformable medical image registration: Setting the state of the art with discrete methods. *Annual review of biomedical engineering* 13, 219–244.

Haber, E., Modersitzki, J., 2007. Intensity gradient based registration and fusion of multi-modal images. *Methods of information in medicine* 46 (3), 292–299.

Heinrich, M., Jenkinson, M., Brady, M., Schnabel, J., 2013a. MRF-based deformable registration and ventilation estimation of lung CT. *IEEE Transactions on Medical Imaging* 32 (7), 1239–1248.

Heinrich, M., Jenkinson, M., Papiez, B., Brady, M., Schnabel, J., 2013b. Edge- and detail-preserving sparse image representations for deformable registration of chest MRI and CT volumes. In: Gee, J., Joshi, S., Pohl, K., Wells, W., L., Z. (Eds.), *Information Processing in Medical Imaging, IPMI 2013*. Lecture Notes in Computer Science. Springer, pp. 463–474.

Heinrich, M. P., Jenkinson, M., Brady, S., Schnabel, J. A., 2011. Non-rigid image registration through efficient discrete optimization. In: *Conference on Medical Image Analysis and Understanding*, pp. 1–5.

Heinrich, M. P., Jenkinson, M., Brady, S. M., Schnabel, J. A., 2012. Globally optimal registration on a minimum spanning tree using dense displacement sampling. In: Delingette, H., Golland, P., Kensaku, M., Ayache, N. (Eds.), *Medical Image Computing and Computer-Assisted Intervention MICCAI 2012*. Lecture Notes in Computer Science. Springer Berlin, pp. 115–122.

Heinrich, M. P., Jenkinson, M., Papiez, B. W., Brady, M., Schnabel, J. A., 2013c. Towards realtime multimodal fusion for image-guided interventions using self-similarities. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2013*. Springer, pp. 187–194.

Heinrich, M. P., Papiez, B., Schnabel, J. A., Handels, H., 2014. Non-parametric discrete registration with convex optimisation. In: Ourselin, S., Modat, M. (Eds.), *Workshop on Biomedical Image Registration, WBIR 2014*. Springer, pp. 53–63.

Heinrich, M. P., Simpson, I. J., Jenkinson, M., Brady, M., Schnabel, J. A., 2013d. Uncertainty estimates for improved accuracy of registration-based segmentation propagation using discrete optimisation. In: *MICCAI Workshop on Segmentation, Algorithms, Theory and Applications*.

Hermann, S., Werner, R., 2013. High accuracy optical flow for 3D medical image registration using the census cost function. In: *PSIVT 2013*. pp. 1–13.

- Iglesias, J. E., Sabuncu, M. R., Van Leemput, K., 2013. Improved inference in bayesian segmentation using monte carlo sampling: Application to hippocampal subfield volumetry. *Medical image analysis* 17 (7), 766–778.
- 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., 2009. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* 46 (3), 786.
- Kohli, P., Torr, P., 2008. Measuring uncertainty in graph cut solutions. *Computer Vision and Image Understanding* 112 (1), 30–38.
- Kolmogorov, V., 2006. Convergent tree-reweighted message passing for energy minimization. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 28 (10), 1568–1583.
- Komodakis, N., Tziritas, G., 2007. Approximate labeling via graph cuts based on linear programming. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29 (8), 1436–1453.
- König, L., Rühaak, J., 2014. A fast and accurate parallel algorithm for non-linear image registration using normalized gradient fields. In: *Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on*. IEEE, pp. 580–583.
- Kopf, J., Cohen, M., Lischinski, D., Uyttendaele, M., 2007. Joint bilateral up-sampling. In: *SIGGRAPH. Proceedings of ACM Conference*.
- Kybic, J., 2010. Bootstrap resampling for image registration uncertainty estimation without ground truth. *Image Processing, IEEE Transactions on* 19 (1), 64–73.
- Lei, C., Yang, Y.-H., 2009. Optical flow estimation on coarse-to-fine region trees using discrete optimization. In: *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, pp. 1562–1569.
- Lempitsky, V., Roth, S., Rother, C., 2008. Fusionflow: Discrete-continuous optimization for optical flow estimation. In: *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, pp. 1–8.
- Lotfi, T., Tang, L., Andrews, S., Hamarneh, G., 2013. Improving probabilistic image registration via reinforcement learning and uncertainty evaluation. In: *Machine Learning in Medical Imaging*. Springer, pp. 187–194.
- Lucchi, A., Smith, K., Achanta, R., Knott, G., Fua, P., 2012. Supervoxel-based segmentation of mitochondria in em image stacks with learned shape features. *Medical Imaging, IEEE Transactions on* 31 (2), 474–486.
- Muenzing, S. E., van Ginneken, B., Murphy, K., Pluim, J. P., 2012a. Supervised quality assessment of medical image registration: Application to intra-patient ct lung registration. *Medical image analysis* 16 (8), 1521–1531.
- Muenzing, S. E., van Ginneken, B., Pluim, J. P., 2012b. On combining algorithms for deformable image registration. In: *Biomedical Image Registration*. Springer, pp. 256–265.
- Muenzing, S. E., van Ginneken, B., Viergever, M. A., Pluim, J. P., 2014. Dirboost—an algorithm for boosting deformable image registration: Application to lung ct intra-subject registration. *Medical Image Analysis*.
- Oreshkin, B., Arbel, T., 2013. Uncertainty driven probabilistic voxel selection for image registration. *Medical Imaging, IEEE Transactions on* 32 (10), 1777–1790.
- Ou, Y., Akbari, H., Bilello, M., Da, X., Davatzikos, C., 2014. Comparative evaluation of registration algorithms in different brain databases with varying difficulty: Results and insights. *Medical Imaging, IEEE Transactions on* 33 (10), 2039–2065.
- Ou, Y., Sotiras, A., Paragios, N., Davatzikos, C., 2011. DRAMMS: deformable registration via attribute matching and mutual-saliency weighting. *Medical image analysis* 15 (4), 622–639.
- Papiez, B., Heinrich, M., Risser, L., Schnabel, J., 2013. Complex lung motion estimation via adaptive bilateral filtering of the deformation field. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2013*. Springer, pp. 25–32.
- Parisot, S., Wells III, W., Chemouny, S., Duffau, H., Paragios, N., 2014. Current tumor segmentation and registration with uncertainty-based sparse non-uniform graphs. *Medical image analysis* 18 (4), 647–659.
- Popuri, K., Cobzas, D., Jägersand, M., 2013. A variational formulation for discrete registration. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (Eds.), *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2013*. Vol. 8151 of Lecture Notes in Computer Science. Springer, pp. 187–194.
- Prim, R. C., 1957. Shortest connection networks and some generalizations. *Bell System Technical Journal* 36, 1389–1401–244.
- Ren, X., Malik, J., 2003. Learning a classification model for segmentation. In: *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*. IEEE, pp. 10–17.
- Risholm, P., Janoos, F., Norton, I., Golby, A. J., Wells III, W. M., 2013. Bayesian characterization of uncertainty in intra-subject non-rigid registration. *Medical image analysis* 17 (5), 538–555.
- Risser, L., Vialard, F., Baluwala, H., Schnabel, J., 2012. Piecewise-diffeomorphic image registration: Application to the motion estimation between 3D CT lung images with sliding conditions. *Medical Image Analysis*.
- Rohde, G. K., Aldroubi, A., Dawant, B. M., 2003. The adaptive bases algorithm for intensity-based nonrigid image registration. *Medical Imaging, IEEE Transactions on* 22 (11), 1470–1479.
- Rousseau, F., Habas, P. A., Studholme, C., 2011. A supervised patch-based approach for human brain labeling. *Medical Imaging, IEEE Transactions on* 30 (10), 1852–1862.
- Rudin, L. I., Osher, S., Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena* 60 (1), 259–268.
- Rueckert, D., Aljabar, P., Heckemann, R. A., Hajnal, J. V., Hammers, A., 2006. Diffeomorphic registration using b-splines. In: Larsen, R., Nielsen, M., Sparring, J. (Eds.), *Medical Image Computing and Computer-Assisted Intervention MICCAI 2006*. Vol. 4190 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg, pp. 702–709.
- Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., Hawkes, D., 1999. Nonrigid registration using free-form deformations: application to breast MR images. *Medical Imaging, IEEE Transactions on* 18 (8), 712–721.
- Schmidt-Richberg, A., Ehrhardt, J., Werner, R., Handels, H., 2012a. Fast explicit diffusion for registration with direction-dependent regularization. *Biomedical Image Registration*, 220–228.
- Schmidt-Richberg, A., Werner, R., Handels, H., J., E., 2012b. Estimation of slipping organ motion by registration with direction-dependent regularization. *Medical Image Analysis* 16 (1), 150–159.
- Schnabel, J., Rueckert, D., Quist, M., Blackall, J., Castellano-Smith, A., Hartkens, T., Penney, G., Hall, W., Liu, H., Truwit, C., Gerritsen, F., Hill, D., Hawkes, D., 2001. A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations. In: Niessen, W., Viergever, M. (Eds.), *Medical Image Computing and Computer-Assisted Intervention, MICCAI 2001*. Vol. 2208 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg, pp. 573–581.
- Scott, D. W., 1979. On optimal and data-based histograms. *Biometrika* 66 (3), 605–610.
- Shattuck, D. W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K. L., Poldrack, R. A., Bilder, R. M., Toga, A. W., 2008. Construction of a 3D probabilistic atlas of human cortical structures. *Neuroimage* 39 (3), 1064.
- Shekhovtsov, A., Kovtun, I., Hlaváč, V., 2008. Efficient MRF deformation model for non-rigid image matching. *Computer Vision and Image Understanding* 112 (1), 91–99.
- Shi, J., Malik, J., 2000. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22 (8), 888–905.
- Shi, W., Jantsch, M., Aljabar, P., Pizarro, L., Bai, W., Wang, H., O’Regan, D., Zhuang, X., Rueckert, D., 2013. Temporal sparse free-form deformations. *Medical image analysis* 17 (7), 779–789.
- Simpson, I. J., Schnabel, J. A., Groves, A. R., Andersson, J. L., Woolrich, M. W., 2012. Probabilistic inference of regularisation in non-rigid registration. *NeuroImage* 59 (3), 2438–2451.
- Simpson, I. J., Woolrich, M. W., Andersson, J. L., Groves, A. R., Schnabel, J. A., 2013. Ensemble learning incorporating uncertain registration. *IEEE Transactions on Medical Imaging* 32 (4), 748–756.
- Simpson, I. J., Woolrich, M. W., Schnabel, J. A., 2011. Probabilistic segmentation propagation from uncertainty in registration. In: *Proceedings of Medical Image Understanding and Analysis (MIUA)*.
- Sotiras, A., Davatzikos, C., N., P., 2013. Deformable medical image registration: A survey. *IEEE Transactions on Medical Imaging* 32 (7), 1153–1190.
- Sotiras, A., Paragios, N., 2012. Discrete symmetric image registration. In: *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*. pp. 342–345.
- Tang, L. Y., Hamarneh, G., 2013. Random walks with efficient search and contextually adapted image similarity for deformable registration. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2013*. Springer Berlin Heidelberg, pp. 43–50.
- Tang, L. Y. W., Hamarneh, G., April 2014. Random walker image registration with cost aggregation. In: *Biomedical Imaging (ISBI), 2014 IEEE 11th In-*

ternational Symposium on. pp. 576–579.

Tarlow, D., Adams, R. P., 2012. Revisiting uncertainty in graph cut solutions. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, pp. 2440–2447.

Tighe, J., Lazebnik, S., 2013. Superparsing. International Journal of Computer Vision 101 (2), 329–349.

Vandemeulebroucke, J., Bernard, O., Rit, S., Kybic, J., Clarysse, P., Sarrut, D., 2012. Automated segmentation of a motion mask to preserve sliding motion in deformable registration of thoracic CT. Medical Physics 39, 1006.

Veksler, O., 2005. Stereo correspondence by dynamic programming on a tree. In: Computer Vision and Pattern Recognition, CVPR 2005. Vol. 2. pp. 384–390.

Wang, H., Suh, J. W., Das, S. R., Pluta, J. B., Craige, C., Yushkevich, P. A., 2013. Multi-atlas segmentation with joint label fusion. Pattern Analysis and Machine Intelligence, IEEE Transactions on 35 (3), 611–623.

Wang, H., Yushkevich, P. A., 2013. Multi-atlas segmentation without registration: A supervoxel-based approach. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013. Springer, pp. 535–542.

Warfield, S. K., Zou, K. H., Wells, W. M., 2004. Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation. Medical Imaging, IEEE Transactions on 23 (7), 903–921.

Wassermann, D., Toews, M., Niethammer, M., Wells III, W., 2014. Probabilistic diffeomorphic registration: Representing uncertainty. In: Biomedical Image Registration. Springer, pp. 72–82.

Weiss, E., Wijesooriya, K., Dill, S. V., Keall, P. J., 2007. Tumor and normal tissue motion in the thorax during respiration: Analysis of volumetric and positional variations using 4D CT. International Journal of Radiation Oncology Biology Physics 67 (1), 296 – 307.

Werner, R., Schmidt-Richberg, A., Handels, H., Ehrhardt, J., 2014. Estimation of lung motion fields in 4d ct data by variational non-linear intensity-based registration: A comparison and evaluation study. Physics in medicine and biology 59 (15), 4247.

Willsky, A. S., 2002. Multiresolution markov models for signal and image processing. Proceedings of the IEEE 90 (8), 1396–1458.

Yang, Q., Wang, L., Ahuja, N., 2010. A constant-space belief propagation algorithm for stereo matching. In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, pp. 1458–1465.

Ye, D. H., Hamm, J., Desjardins, B., Pohl, K. M., 2013. Floor: Fusing locally optimal registrations. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013. Springer, pp. 195–202.

Yushkevich, P. A., Wang, H., Pluta, J., Avants, B. B., 2012. From label fusion to correspondence fusion: a new approach to unbiased groupwise registration. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, pp. 956–963.

Zitnick, C. L., Kang, S. B., 2007. Stereo for image-based rendering using image over-segmentation. International Journal of Computer Vision 75 (1), 49–65.

DIR-lab case		avg	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
A	$\mu(\text{TRE})$	2.821	1.94	2.08	1.93	3.18	2.82	2.94	3.72	4.79	2.05	2.76
	neg. J	5.88E-03	1E-03	3E-03	5E-03	5E-03	5E-03	1E-02	7E-03	1E-02	3E-03	1E-02
B	$\sigma(\text{Jac})$	0.292	0.23	0.24	0.29	0.32	0.30	0.30	0.29	0.36	0.26	0.34
	$\mu(\text{Jac})$	0.887	0.95	0.94	0.93	0.90	0.96	0.79	0.82	0.84	0.87	0.86
	harm. e.	0.395	0.22	0.29	0.52	0.45	0.45	0.39	0.31	0.57	0.27	0.48
	$\mu(\text{TRE})$	1.619	1.71	1.38	1.72	1.78	1.72	1.61	1.54	1.66	1.49	1.57
C	neg. J	4.03E-03	1E-03	4E-03	5E-03	3E-03	5E-03	6E-03	3E-03	8E-03	8E-04	8E-03
	$\sigma(\text{Jac})$	0.251	0.19	0.21	0.26	0.27	0.29	0.23	0.22	0.35	0.19	0.30
	$\mu(\text{Jac})$	0.889	0.95	0.94	0.93	0.90	0.96	0.80	0.82	0.84	0.88	0.87
	harm. e.	0.333	0.16	0.25	0.46	0.36	0.45	0.28	0.19	0.56	0.15	0.46
	$\mu(\text{TRE})$	1.326	1.27	1.10	1.27	1.46	1.51	1.33	1.31	1.43	1.28	1.29
	neg. J	1.95E-05	0E+00	4E-06	0E+00	0E+00	6E-06	2E-05	2E-06	1E-04	0E+00	1E-04
D	$\sigma(\text{Jac})$	0.126	0.09	0.10	0.12	0.16	0.13	0.12	0.12	0.14	0.12	0.14
	$\mu(\text{Jac})$	0.880	0.95	0.94	0.93	0.90	0.96	0.78	0.81	0.83	0.86	0.85
	harm. e.	0.157	0.06	0.11	0.18	0.16	0.17	0.16	0.12	0.27	0.10	0.23
	$\mu(\text{TRE})$	1.236	1.19	1.01	1.20	1.36	1.42	1.22	1.27	1.26	1.20	1.23
	$\sigma(\text{TRE})$	0.816	0.64	0.62	0.68	0.96	1.32	0.70	0.71	0.95	0.71	0.88
	neg. J	2.92E-05	0E+00	0E+00	0E+00	0E+00	4E-06	3E-06	4E-05	2E-04	0E+00	9E-05
E	$\sigma(\text{Jac})$	0.126	0.09	0.10	0.11	0.16	0.13	0.13	0.13	0.15	0.12	0.15
	$\mu(\text{Jac})$	0.881	0.95	0.94	0.93	0.90	0.96	0.78	0.81	0.83	0.86	0.85
	harm. e.	0.134	0.04	0.07	0.12	0.12	0.11	0.17	0.12	0.28	0.10	0.22
	$\mu(\text{TRE})$	1.257	1.19	1.01	1.20	1.37	1.48	1.24	1.29	1.26	1.24	1.28

Table 3: Evaluation metrics for each case of DIR-lab lung dataset. All four variants of our proposed algorithm are compared in terms of accuracy (TRE) and complexity of transformations.

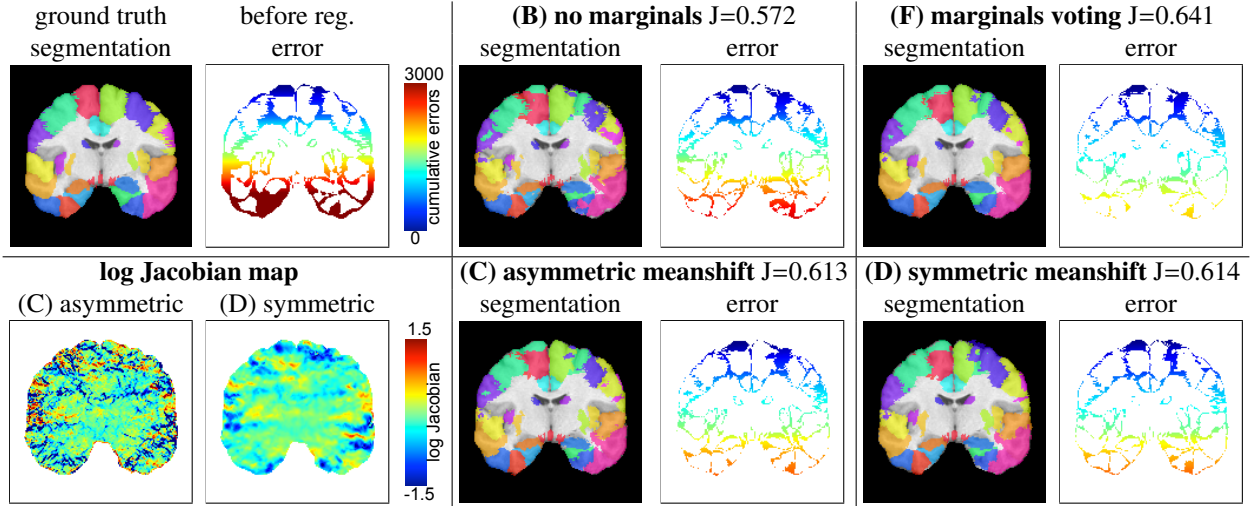


Figure 9: Visual examples for LPBA40 dataset for registration and segmentation propagation showing the **best** case (#13 and #31).

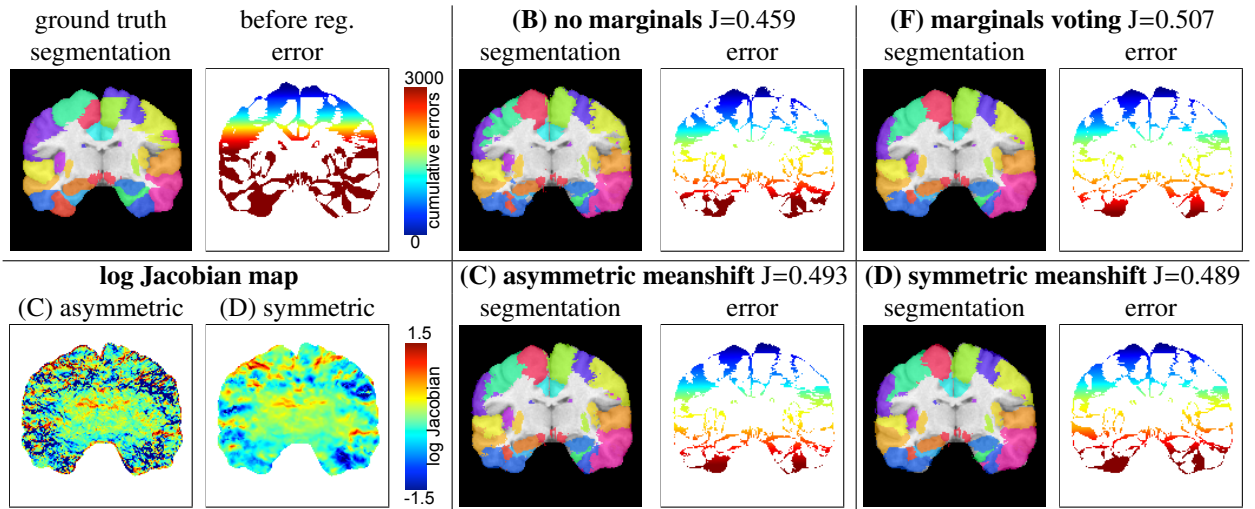


Figure 10: Visual examples for LPBA40 dataset for registration and segmentation propagation showing the **worst** case (#19 and #34).

Region	ART	IRTK	SyN	variant B	variant C	variant D	variant E	variant F	DICE F
L sup. frontal gyrus	0.690	0.660	0.690	0.661	0.693	0.688	0.694	0.713	0.832
R sup. frontal gyrus	0.680	0.660	0.680	0.653	0.683	0.679	0.684	0.705	0.826
L middle frontal gyrus	0.650	0.630	0.660	0.630	0.659	0.655	0.660	0.680	0.808
R middle frontal gyrus	0.640	0.630	0.650	0.616	0.644	0.641	0.646	0.665	0.797
L inf. frontal gyrus	0.580	0.570	0.570	0.543	0.576	0.577	0.579	0.596	0.744
R inf. frontal gyrus	0.570	0.560	0.560	0.529	0.555	0.561	0.564	0.578	0.729
L precentral gyrus	0.580	0.570	0.590	0.532	0.585	0.578	0.582	0.606	0.750
R precentral gyrus	0.560	0.550	0.580	0.518	0.568	0.567	0.574	0.595	0.743
L middle orbitofrontal gyrus	0.530	0.510	0.520	0.500	0.535	0.526	0.532	0.553	0.706
R middle orbitofrontal gyrus	0.530	0.500	0.520	0.501	0.534	0.527	0.530	0.551	0.706
L lateral orbitofrontal gyrus	0.440	0.430	0.430	0.419	0.454	0.444	0.449	0.472	0.636
R lateral orbitofrontal gyrus	0.420	0.400	0.420	0.391	0.420	0.414	0.419	0.439	0.602
L gyrus rectus	0.520	0.490	0.520	0.492	0.536	0.527	0.531	0.552	0.708
R gyrus rectus	0.540	0.500	0.540	0.515	0.554	0.545	0.550	0.576	0.728
L postcentral gyrus	0.510	0.490	0.520	0.462	0.509	0.503	0.502	0.532	0.685
R postcentral gyrus	0.500	0.490	0.520	0.463	0.518	0.517	0.519	0.542	0.698
L sup. parietal gyrus	0.560	0.540	0.580	0.523	0.557	0.555	0.559	0.589	0.738
R sup. parietal gyrus	0.560	0.540	0.570	0.517	0.551	0.552	0.554	0.580	0.732
L supramarginal gyrus	0.490	0.480	0.510	0.452	0.486	0.489	0.478	0.510	0.669
R supramarginal gyrus	0.470	0.480	0.490	0.450	0.486	0.488	0.483	0.510	0.669
L angular gyrus	0.460	0.450	0.470	0.432	0.461	0.460	0.458	0.485	0.647
R angular gyrus	0.480	0.480	0.490	0.455	0.486	0.485	0.477	0.509	0.671
L precuneus	0.510	0.480	0.510	0.454	0.499	0.505	0.512	0.526	0.686
R precuneus	0.520	0.490	0.520	0.473	0.521	0.524	0.526	0.543	0.701
L sup. occipital gyrus	0.430	0.410	0.430	0.393	0.432	0.434	0.435	0.454	0.617
R sup. occipital gyrus	0.410	0.390	0.410	0.374	0.403	0.402	0.401	0.424	0.587
L middle occipital gyrus	0.520	0.490	0.520	0.490	0.521	0.524	0.523	0.541	0.699
R middle occipital gyrus	0.510	0.490	0.510	0.489	0.518	0.521	0.518	0.541	0.699
L inf. occipital gyrus	0.510	0.450	0.510	0.470	0.507	0.500	0.505	0.524	0.683
R inf. occipital gyrus	0.530	0.460	0.520	0.482	0.522	0.511	0.515	0.538	0.696
L cuneus	0.500	0.450	0.500	0.440	0.500	0.497	0.506	0.520	0.678
R cuneus	0.480	0.440	0.490	0.432	0.485	0.478	0.479	0.501	0.660
L sup. temporal gyrus	0.610	0.620	0.630	0.585	0.621	0.606	0.611	0.648	0.785
R sup. temporal gyrus	0.620	0.620	0.640	0.608	0.640	0.624	0.633	0.671	0.802
L middle temporal gyrus	0.510	0.500	0.510	0.482	0.511	0.510	0.510	0.541	0.699
R middle temporal gyrus	0.530	0.530	0.540	0.510	0.542	0.541	0.545	0.570	0.724
L inf. temporal gyrus	0.510	0.490	0.510	0.482	0.513	0.508	0.516	0.535	0.694
R inf. temporal gyrus	0.520	0.510	0.530	0.499	0.530	0.525	0.533	0.553	0.710
L parahippocampal gyrus	0.560	0.540	0.560	0.543	0.581	0.563	0.567	0.613	0.758
R parahippocampal gyrus	0.550	0.520	0.540	0.534	0.566	0.547	0.554	0.598	0.746
L lingual gyrus	0.560	0.510	0.570	0.522	0.561	0.550	0.563	0.584	0.733
R lingual gyrus	0.580	0.540	0.590	0.541	0.580	0.566	0.572	0.600	0.746
L fusiform gyrus	0.560	0.510	0.550	0.515	0.556	0.548	0.550	0.582	0.733
R fusiform gyrus	0.560	0.510	0.550	0.533	0.569	0.558	0.564	0.595	0.743
L insular cortex	0.670	0.680	0.670	0.637	0.664	0.664	0.674	0.691	0.817
R insular cortex	0.650	0.650	0.640	0.610	0.637	0.636	0.650	0.662	0.795
L cingulate gyrus	0.550	0.540	0.550	0.502	0.540	0.542	0.553	0.556	0.712
R cingulate gyrus	0.540	0.530	0.540	0.497	0.530	0.532	0.540	0.548	0.703
L caudate	0.610	0.630	0.610	0.561	0.586	0.585	0.593	0.600	0.745
R caudate	0.610	0.620	0.600	0.568	0.590	0.591	0.596	0.607	0.751
L putamen	0.660	0.640	0.640	0.621	0.651	0.651	0.655	0.670	0.802
R putamen	0.650	0.630	0.640	0.616	0.649	0.650	0.653	0.668	0.800
L hippocampus	0.620	0.600	0.610	0.582	0.617	0.621	0.626	0.641	0.780
R hippocampus	0.620	0.600	0.610	0.586	0.617	0.620	0.625	0.640	0.779
cerebellum	0.760	0.730	0.750	0.721	0.760	0.763	0.756	0.767	0.867
brainstem	0.680	0.670	0.660	0.606	0.679	0.694	0.691	0.681	0.808
mean	0.557	0.537	0.558	0.522	0.558	0.555	0.558	0.580	0.728
stddev	0.074	0.078	0.072	0.074	0.073	0.073	0.074	0.072	0.059

Table 4: Detailed results for segmentation accuracy (Jaccard overlap unless otherwise stated) for all regions in LPBA40 dataset. All variants of the presented method are compared to the three best performing algorithms ART, IRTK and SyN from the study of (Klein et al., 2009) as published at: http://www.mindboggle.info/papers/evaluation_NeuroImage2009/tables/table_ROIMethod_UO_LPBA40.html.