

Globally Optimal Deformable Registration on a Minimum Spanning Tree using Dense Displacement Sampling

Mattias P. Heinrich^{1,2,*}, Mark Jenkinson², Sir Michael Brady³, and Julia A. Schnabel¹

¹ Institute of Biomedical Engineering, University of Oxford, UK

² Oxford University Centre for Functional MRI of the Brain, UK

³ Department of Oncology, University of Oxford, UK

mattias.heinrich@eng.ox.ac.uk, <http://users.ox.ac.uk/~shil3388>

Abstract. Deformable image registration poses a highly non-convex optimisation problem. Conventionally, medical image registration techniques rely on continuous optimisation, which is prone to local minima. Recent advances in the mathematics and new programming methods enable these disadvantages to be overcome using discrete optimisation. In this paper, we present a new technique **deeds**, which employs a **discrete dense displacement sampling** for the deformable registration of high resolution CT volumes. The image grid is represented as **a minimum spanning tree**. Given these constraints a global optimum of the cost function can be found efficiently **using dynamic programming**, which enforces the smoothness of the deformations. Experimental results demonstrate the advantages of **deeds**: the registration error for the challenging registration of inhale and exhale pulmonary CT scans is significantly lower than for two state-of-the-art registration techniques, **especially** in the presence of large deformations and sliding motion at lung surfaces.

1 Introduction

Deformable image registration is a key enabling technique in medical image analysis. Applications include motion correction, image-guided radiotherapy, multi-modality fusion, quantification of longitudinal progression of disease, and inter-subject registration for atlas-based segmentation. **Non-rigid registration algorithms aim to solve a highly non-convex optimisation problem with several million degrees of freedom**. In the medical domain, this problem has so far been addressed almost exclusively using continuous optimisation. In a recent comparison study of non-rigid registration methods applied to pulmonary CT scans [1], 23 out of 24 algorithms used continuous optimisation. However, continuous optimisation of a non-convex cost function is susceptible to local minima, potentially leading to an erroneous registration. Local minima are frequently encountered

* We thank EPSRC and Cancer Research UK for funding this work in the Oxford Cancer Imaging Centre. JAS also acknowledges funding from EPSRC EP/H050892/1.

when the initial motion estimate is far from the desired solution, especially for small anatomical features, which have undergone large deformations. A second drawback for continuous optimisation is the necessity either for an analytical or for a numerical derivative of the cost function, thus limiting its choice.

Discrete optimisation offers numerous advantages to overcome these limitations. It has been widely used for 2D computer vision applications. However, in discrete optimisation the space of possible displacements \mathcal{L} has to be quantised. If the smallest discretisation step is defined to be one voxel, the displacement space can be of the order of thousands and therefore there are millions of degrees of freedoms. This high computational and memory demands have previously discouraged its use in 3D problems. One state-of-the-art medical image registration method, based on discrete optimisation, **drop** [2], reduces the dimensionality of the problem by using a parametric transformation model based on a B-spline deformation grid. Additionally, the displacement space is sampled only sparsely (along the three axes). This may possibly cause the optimal displacement to be missed. The authors attempt to address this problem, first by introducing a multi-resolution scheme, and second by iteratively updating the transformation (thus warping the source image towards the target). While the method is an improvement (both in terms of accuracy and speed) over commonly used continuous optimisation counterparts, it lacks some of the attractive properties that a discrete framework potentially offers: most notably, avoidance of an iterative solution and image interpolation; and the use of a dense sampling of the displacement space.

In this work, we address these shortcomings in three ways. First, we reformulate the image grid so that it is not fully connected. Instead, a minimum spanning tree [3] is computed, which best replicates the underlying structure of the anatomical connectivity of the image. This allows us to use dynamic programming to find the global optimum of the registration cost non-iteratively in just two passes. Second, we use the min-convolution technique [4] to reduce the complexity of the pair-wise regularisation cost computation from $\mathcal{O}(|\mathcal{L}|^2)$ to $\mathcal{O}(|\mathcal{L}|)$. Finally, we use a multi-level approach in which groups of voxels are represented by a single node in the graph. This leads to an efficient coarse-to-fine optimisation, while, at the same time, all data terms are calculated in the original image resolution (so there is no image degradation as a result of downsampling). At subsequent (finer) levels, the previous solution is used as a prior of the displacement. The improved computational efficiency allows us to use a very large label space and thus better address the efficiency-quality trade-off than previous approaches. We employ a symmetric diffeomorphic approach to ensure unbiased physically plausible transformations. Our approach is called **deeds** (dense displacement sampling) and is explained in detail in the next section. An experimental validation is performed for the non-rigid registration of pulmonary CT scans at different breathing states. We compare our approach to a continuous optimisation algorithm (**gsyn** [5]), which performed best in the recent EMPIRE study [1]; and a discrete parametric optimisation framework (**drop**) using linear programming [2].

2 Deformable Registration using deeds

Discrete optimisation is usually performed as Markov Random Field (MRF) labelling. For the purposes of our non-parametric image registration framework, a graph is defined, in which the nodes $p \in \mathcal{P}$ correspond to voxels (or groups of voxels) and in which, for each node, there is a set of (hidden) labels f_p , which correspond to discrete displacements. The energy function to be optimised consists of two terms: the data (also called unary) cost D (which is independent for each node); and the pair-wise regularisation cost $R(f_p, f_q)$ for any node q , which is directly connected ($\in \mathcal{N}$) with p :

$$E(f) = \underbrace{\sum_{p \in \mathcal{P}} D(f_p)}_{\text{data term}} + \alpha \underbrace{\sum_{(p,q) \in \mathcal{N}} R(f_p, f_q)}_{\text{regularisation term}} \quad (1)$$

The unary cost measures the similarity of a voxel in one image and a displaced voxel in the second image, and is independent of the displacements of its neighbours. The pair-wise term enforces a globally smooth transformation by penalising deviations of the displacements of neighbouring voxels. The weighting parameter α sets the influence of the regularisation.

Methods to solve the MRF labelling problem can generally be categorised as one of two approaches: message passing and graph cuts. Message passing schemes include: loopy belief propagation (LBP); sequential tree-reweighted message passing (TRW-S); and dynamic programming on a tree. Popular graph cut algorithms include: α -expansion moves; and the fast primal-dual strategy (FastPD). For some problems, graph cuts guarantee convergence to a known bound, close to the global optimum. However, in practice, the complexity limits both the number of nodes and the label space (directly applied to medical images α -expansion graph cuts take up to 24 hours for a single registration [6]). In contrast, our approach can find the global optimum for a complex 3D registration problem with a very large (dense) label space within minutes using a reduced neighbourhood interaction based on a minimum spanning tree (MST).

2.1 Dynamic Programming on Minimum Spanning Tree (MST)

Optimising the registration cost function on a six-connected graph is NP-hard. In contrast, the very efficient dynamic programming technique finds a global minimum, without iterations, in just two passes on a cycle-free graph (e.g. a tree). Using Prim's algorithm [7], we can quickly find the unique MST given a set of nodes $p \in \mathcal{P}$ and edges e . The edge weight $w(p, q)$ is defined as the sum of absolute differences (SAD) between the intensities of all voxels contained within the node p and the respective voxels in node q . The MST is a spanning tree with minimum total edge costs. The selection of the root node does not influence the result of the optimisation. The MST sufficiently reflects the underlying anatomical connectivity in a medical image (see Fig. 1 left). It is well balanced, and, as a consequence, the maximum width is approximately $|\mathcal{P}|/\log|\mathcal{P}|$. The output

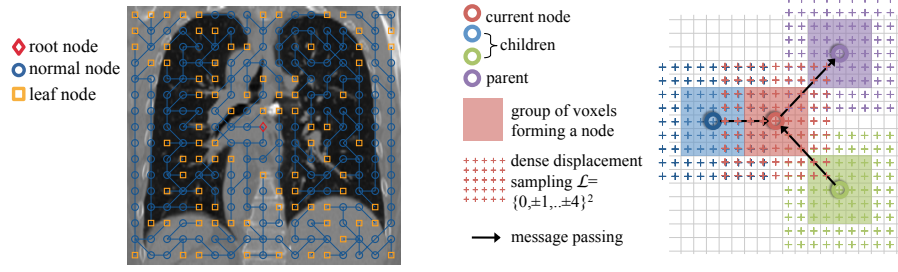


Fig. 1: An example of minimum spanning tree (MST) of a 2D coronal slice of a lung CT is shown on the left (\diamond root, \circ normal, and \square leaf nodes). The concept of our proposed discrete optimisation scheme, using dense displacement sampling, is displayed on the right. Note that, even though the nodes may be sparsely distributed, the data cost is always computed in the original resolution.

of Prim’s algorithm consists of a sorted list of all nodes (with increasing tree depth) and the index of each node’s parent. A similar approach has been used for stereo correspondence [3] however, other methods (LBP, TRW-S) perform better on that specific application.

Finding the best labelling, and thereby the global optimum of Eq. 1, is possible using dynamic programming on the MST [3]. At each node p , the cost C_p of the best displacement can be found, given the displacement f_q of its parent q :

$$C_p(f_q) = \min_{f_p} \left(D(f_p) + R(f_p, f_q) + \sum_c C_c(f_p) \right) \quad (2)$$

where c are the children of p . The best displacement can be found by replacing min with argmin in Eq. 2. For any leaf node, Eq. 2 can be evaluated directly (since it has no children). **Thereafter, the tree is traversed from its leaves down to the root node.** It is worth noting that the costs C_p only have to be stored for the next tree level (only the argmin is needed to select the best displacement). Once the root node is reached, the best labelling for all nodes can be selected in another pass through the tree (from root to leaves). Finding the minimum naïvely would require $|\mathcal{L}|^2$ calculations for the regularisation cost per node. In [4] the min-convolution technique is introduced, which reduces the complexity to $|\mathcal{L}|$ by employing a lower envelope computation. For most commonly used (pair-wise) regularisation terms, such as diffusion (squared difference of displacements) and total variation (absolute difference) regularisation, this simplification is possible.

2.2 Dense displacement sampling

To avoid local minima, most continuous-optimisation-based registration algorithms use a multi-resolution scheme in which the images are downsampled after a prior Gaussian smoothing. This may degrade the quality of the registration.

We adopt a different approach: a multi-level scheme, in which we only employ the highest resolution image. For a given level, the image is subdivided into non-overlapping cubic groups of voxels. The similarity cost is first calculated for each voxel separately using dense displacement sampling, then aggregated for all voxels of the same group (this forms an additional intrinsic regularisation and reduces the number of nodes). Subsequently, the regularisation term is calculated only for each group of voxels (see Fig. 1). Using this approach, both high spatial accuracy and low computational complexity are achieved. The optimal labelling is obtained as explained previously. For the next level, the previous deformation field is upsampled and used as the prior deformation. At a finer level, the sampling range may be reduced, because the optimal transformation for the coarser level is known and only a small deviation from it is expected.

2.3 Symmetric and diffeomorphic transformations

For many deformable registration algorithms, one image has to be chosen as the target, the other as the moving image. This biases the registration outcome and may additionally introduce an inverse consistency error (ICE). The ICE for a forward transform A and a backward transform B is defined as the difference between AB and the identity. **gsyn** [5] uses a symmetric deformable registration, which calculates a transformation from both images to a common intermediate image and also ensures that $A(0.5) = B(0.5)^{-1}$. The full forward transformation is then $A(0.5) \circ B(0.5)^{-1}$. We adopt a similar approach and estimate both A and B . We then use a fast iterative inversion method, as presented in [8], to obtain $A(0.5)^{-1}$ and $B(0.5)^{-1}$. Additionally, if we treat the displacement field as a velocity field, we can transform it into a diffeomorphic mapping by applying the scaling and squaring method [9]. This approach avoids transformations for which physically implausible folding of volume occurs. Although this yields a continuous-valued transformation, we need to apply the voxel-sized discretisation for the next level. However, the discretisation is not performed for the final level.

3 Experiments and Results

We evaluate our deformable registration method on ten cancer (esophagus or lung) patients of the DIR-Lab 4D CT dataset⁴ between the inhale and exhale phases (no intermediate frames were used in our experiments). The scans were acquired as a breathing cycle CT of the thorax and upper abdomen with a spatial resolution of 0.97 to 1.16 mm in the xy-direction and 2.5 mm in the z-direction. These registration tasks are particularly challenging due to: the large deformations of small features (lung vessels, airways); the discontinuous sliding motion between lung lobes and the lung rib cage interface; and changing contrast due to the compression of air. For each pair, 300 anatomical landmarks were selected by a clinical expert, with an inter-observer error of less than 1 mm [10]. We compare our method to **gsyn** [5], which is a symmetric, diffeomorphic,

⁴ The 4D CT dataset with landmarks is freely available at <http://www.dir-lab.com>

Table 1: Results for deformable registration of inhale and exhale CT. Average target registration error (TRE) for 300 expert selected landmarks per scan pair is given in mm. Best performing algorithm per case is set in bold (for comparison: \emptyset TRE obtained by [11] for the same dataset is 2.13 ± 1.8 mm). Average computation time, (maximum) degrees of freedom (d.o.f.), and harmonic energy of the deformation fields. The deformation fields of both **deeds** and **gsyn** have no negative Jacobians, the ones of **drop** exhibit a small fraction of 8.5×10^{-4} .

| # | initial | drop | gsyn | deeds | # | initial | drop | gsyn | deeds |
|----------|----------------|----------------|----------------------------------|----------------------------------|--------------|-------------------|-------------------|-------------------|----------------------------------|
| 1 | 3.89 ± 2.8 | 1.10 ± 0.7 | 1.03 ± 0.5 | 0.80 ± 0.7 | 8 | 15.0 ± 9.0 | 6.64 ± 8.2 | 7.06 ± 8.6 | 2.78 ± 3.1 |
| 2 | 4.34 ± 3.9 | 1.12 ± 0.8 | 1.02 ± 0.6 | 0.86 ± 0.7 | 9 | 7.92 ± 4.0 | 2.89 ± 2.6 | 1.89 ± 2.0 | 1.35 ± 0.8 |
| 3 | 6.94 ± 4.1 | 1.74 ± 1.4 | 1.31 ± 1.0 | 1.14 ± 0.8 | 10 | 7.30 ± 6.3 | 2.03 ± 2.6 | 2.06 ± 3.0 | 1.50 ± 1.4 |
| 4 | 9.83 ± 4.9 | 2.78 ± 4.1 | 1.65 ± 1.6 | 1.71 ± 1.7 | \emptyset | 8.46 ± 6.6 | 2.85 ± 4.0 | 2.43 ± 4.1 | 1.60 ± 1.7 |
| 5 | 7.48 ± 5.5 | 2.03 ± 2.0 | 2.05 ± 2.3 | 1.77 ± 1.7 | avg. time | 8 min | 29 min | 20 min | |
| 6 | 10.9 ± 7.0 | 5.20 ± 4.5 | 2.50 ± 3.3 | 1.88 ± 1.4 | d.o.f. | 3.7×10^6 | 2.2×10^7 | 9.8×10^8 | |
| 7 | 11.0 ± 7.4 | 3.02 ± 3.4 | 3.77 ± 5.7 | 2.21 ± 2.3 | harm. energy | 0.12 | 0.05 | 0.13 | |

continuous optimisation-based registration tool, and **drop** [2], which is a discrete optimisation method using a B-spline deformation grid. In order to have a fair comparison, we used the same parameter settings as those chosen by the authors themselves in the EMPIRE challenge. For all approaches, 4 levels were used. As similarity metric, SAD is used for **drop** and **deeds**, and normalized cross correlation (NCC) with a radius of 2 voxels for **gsyn** (SAD is not differentiable). The dense sampling range for **deeds** is defined to be $\mathcal{L} = \{0, \pm 1, \dots, \pm 15\}^3$ voxels, ($|\mathcal{L}| = 29791$) for the coarsest level (for datasets (1-5), with smaller deformations, the sampling range of **deeds** was decreased to $\mathcal{L}_{\max} = 10$). At each subsequent level, the range is halved. In our multi-level framework, we use cubic groups of voxels of sizes 6^3 , 4^3 and 2^3 . For **drop** the memory requirements (using 3.5 GB of RAM) limit us to use a sparse sampling of $\mathcal{L} = 3 \times \{0, \pm 1, \dots, \pm 10\}$ ($|\mathcal{L}| = 61$), with a range of 24 mm. The Gaussian smoothing for gradient and deformation fields in **gsyn** is set to 3 and 1 voxels respectively. The regularisation parameters in **drop** $\lambda = 5$ and **deeds** $\alpha = 50$ were empirically chosen (intensities are in the range $[0, 256]$); a quadratic penalty function is used for both methods.

The resulting target registration error (TRE, see Table 1) of **deeds** is 1.25 mm lower than **drop**, and 0.83 mm lower than **gsyn**. For cases with larger deformations the differences are most substantial, because the larger search space (and degrees of freedom, see Table 1) of **deeds** helps to avoid local minima. The improvements are significant, based on a Wilcoxon rank test ($p < 3 \times 10^{-4}$). The registration outcome for Case 9 is illustrated in Fig. 2, where an improved alignment using our method can clearly be seen. Another advantage of the non-iterative dense displacement sampling is the preservation of naturally occurring discontinuous motion fields. This sliding motion occurs when the lung lobes slip along their surfaces at the boundary to the rib cage and between fissures. Figure 2 gives an example, where this sliding motion is clearly preserved using **deeds** while the registration using **drop** or **gsyn** smoothes over the motion boundary.

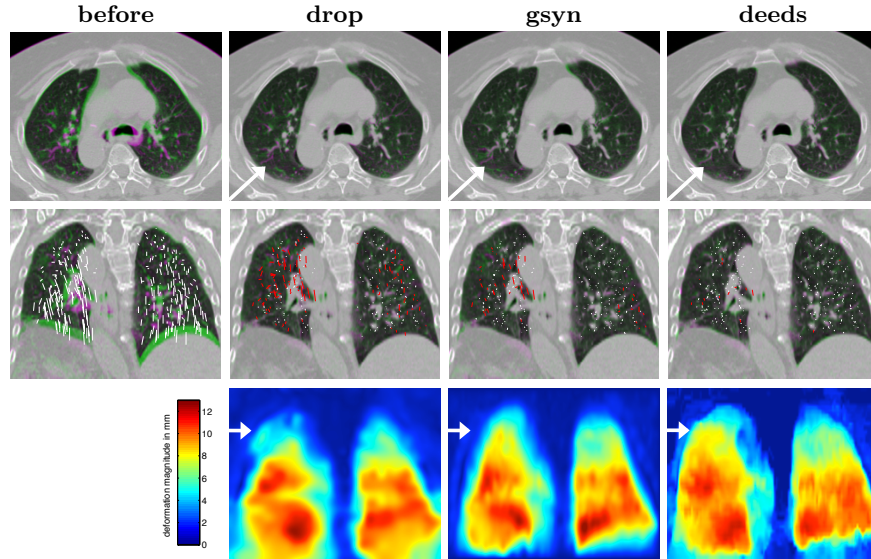


Fig. 2: Registration result for Case 9 of 4D CT dataset. The target volume is displayed in magenta, the moving volume in green (complementary colour). Axial slices before and after registration are shown in the top row. The arrow points to an improved alignment of lung vessels using our approach. The second row shows the coronal plane along with vectors indicating the registration error (errors larger than the voxel size are marked in red). The magnitudes of deformation fields is shown in the bottom row. The sliding motion of the lungs against rib cage and the heart (descending aorta) is better preserved using **deeds**. The deformation fields of **drop** and **gsyn** are too smooth close to the motion boundary at the lung surface (see arrow).

4 Conclusion and Discussion

We have introduced a novel deformable registration method **deeds** that uses discrete optimisation. A **dense displacement sampling** is performed for the similarity term on the highest available resolution. The regularity of the deformation field is obtained using **dynamic programming** on a **minimum spanning tree**. This ensures a globally optimal solution without the need for an iterative scheme. The algorithm is efficiently implemented in a **symmetric multi-level framework** yielding comparable computation time to state-of-the-art algorithms, but employing many more degrees of freedom. An average TRE of 1.60 mm was found for a challenging dataset of inhale-exhale CT scans. This is a significant improvement over the most popular discrete optimisation framework **drop** [2] (TRE=2.85 mm) and **gsyn** [5] (TRE=2.43 mm), which performed best on the recent lung registration challenge (EMPIRE). The TRE was found to be higher than the results from the EMPIRE challenge [1] for several reasons (larger deformations,

lower resolution, and only one annotation per landmark). Most importantly, in contrast to the EMPIRE study, no lung segmentations were used to guide the registration (and mask out intensities outside the lungs). Lung masks can substantially increase the registration accuracy within the lungs. However, this introduces an additional step and does not recover the full physical deformation. Our registration errors (using only inhale and exhale volumes) are smaller than the voxel size for all cases and comparable to the best performing technique on this dataset (4DLTM), which utilises all frames of the breathing cycle. In the presence of large deformations, our proposed dense displacement sampling yields a higher robustness against misregistration than **drop** and **gsyn**. Furthermore, it intrinsically deals very well with the sliding motion at the pleural interface, and hence avoids numerically more complex modelling as done e.g. in [11]. In the future, we plan to apply this new optimisation method on multi-modal datasets. Here we believe the improvements will be even more significant, due to the higher ambiguity in the similarity term.

References

1. Murphy, K., van Ginneken, B., Reinhardt, J., Kabus, S., Ding, K., Deng, X., Pluim, J.: Evaluation of registration methods on thoracic CT: The EMPIRE10 challenge. *IEEE Trans. Med. Imag.* **30**(11) (2011) 1901–1920
2. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registrations through MRFs and efficient linear programming. *Med. Imag. Anal.* **12**(6) (2008) 731–741
3. Veksler, O.: Stereo correspondence by dynamic programming on a tree. In: *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society (2005) 384–390
4. Felzenszwalb, P., Huttenlocher, D.: Efficient belief propagation for early vision. *Int. J. Comp. Vis.* **70**(1) (2006) 41–54
5. Avants, B., Epstein, C., M., G., Gee, J.: Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Med. Imag. Anal.* **12**(1) (2008) 26–41
6. So, R., Tang, T., Chung, A.: Non-rigid image registration of brain magnetic resonance images using graph-cuts. *Pattern Recogn.* **44**(10) (2011) 2450–2467
7. Prim, R.: Shortest connection networks and some generalizations. *Bell System Technical Journal* **36** (1957) 1389–1401
8. Chen, M., Lu, W., C., Q., Ruchala, K., Olivera, G.: A simple fixed-point approach to invert a deformation field. *Medical Physics* **35** (2007) 81
9. Arsigny, V., Commowick, O., Pennec, X., Ayache, N.: A log-euclidean framework for statistics on diffeomorphisms. In Larsen, R., M., N., Sparring, J., eds.: *MICCAI 2006*. Volume 4190 of LNCS., Springer, Heidelberg (2006) 924–931
10. Castillo, R., Castillo, E., Guerra, R., Johnson, V., McPhail, T., Garg, A., Guerrero, T.: A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Phy. Med. Biol.* **54**(7) (2009) 1849
11. Schmidt-Richberg, A., Werner, R., Handels, H., Ehrhardt, J.: Estimation of slipping organ motion by registration with direction-dependent regularization. *Med. Imag. Anal.* **16**(1) (2012) 150–159