Dense Displacement Sampling and Self-Similarities applied to the EMPIRE10 Dataset

Mattias P. Heinrich[1,2,⋆], Mark Jenkinson[2], Sir Michael Brady[3], and Julia A. Schnabel[1]

[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. The EMPIRE10 dataset [8] is a publicly available dataset to evaluate and compare state-of-the-art registration algorithms. Up to date, most successful, highly ranked algorithms make explicit use of the provided semi-automatic lung segmentation masks to deal with the complex problem of sliding motion. In this paper, we extend our previous work on 4D-CT lung registration with implicit sliding motion modelling [3], which does not require any anatomical information. Our approach, which uses an image-derived minimum-spanning-tree to model piece-wise smooth motion fields, provides accurate motion estimation for the whole image domain. To deal with the very challenging cases of inspiration to expiration registration, we include the use of self-similarity descriptors [7] as data term in our discrete optimisation framework. An efficient CPU implementation of our method is publicly available, which achieves computation times of only three minutes per full registration.

1 Introduction and Background

Deformable lung registration has various important clinical applications, such as the propagation of segmented regions of interest across different time points, longitudinal monitoring of lung disease, motion compensation for radiotherapy treatment delivery and motion modelling for treatment planning. Three major challenges are commonly encountered when estimating motion between lung CT scans:

1. Large motion of small anatomical features
2. Sliding motion at the surface of the lung pleura
3. Locally varying contrast due to lung compression

The first challenge mainly concerns the optimisation of the registration cost function, where continuous optimisation strategies often get trapped in local minima or require too many iterations to converge. In [3] and [6], we addressed this problem by introducing a **de**ns**e** **d**isplacement **s**ampling (**deeds**) in combination

⋆ We thank EPSRC and Cancer Research UK for funding this work.

with a discrete optimisation strategy based on belief propagation, which will be explained in detail in Sec. 2.2.

Dealing explicitly with the challenging problem of sliding motion has been finessed by most participants in the EMPIRE10 challenge. Most algorithms use the provided semi-automatic lung segmentation masks to remove all outer-lung content from the scans and only align the inner-lung features. For clinical practice, the requirement of an additional semi-automatic segmentation step would delay the processing pipeline and hinder the deployment of many approaches. Furthermore, masking of certain image regions could have a negative impact for ventilation estimation and tumour motion prediction, especially in cases where the lung segmentation is more challenging due to the pathology (e.g. a lung tumour attached to the pleura). In [3] and [6], we introduced a new graph-based deformation model, which restricts the regularisation term to act only on connections (edges) between control points that have a similar appearance (see Sec. 2.2).

Finding correct anatomical correspondences is another important aspect of lung registration, which is difficult because of locally varying contrast due to compression of the lungs during breathing. Building upon our previous work on multi-modal image registration [4, 5, 7], we use the self-similarity context descriptor as a robust and very discriminative similarity metric, which can be very efficiently calculated as detailed in 2.1. Implementation details and parameter choices are described in Sec. 3, the source code will be made available to enable an exact replication of our method. The results are discussed in Sec. 4.

2 Methods

Our approach addresses the aforementioned challenges of deformable lung registration with a number of contributions. First, a similarity term based on local self-similarity patterns is introduced in Sec. 2.1 [5, 7], which is able to cope with changes in contrast, image noise and difficult anatomical correspondences. Second, a dense displacement sampling [3, 6] with a very large space of potential discrete deformations is performed to deal with the large motion of small anatomical structures, which are particularly problematic for inhale and exhale scan pairs. An efficient discrete optimisation framework is used to infer an approximately globally optimal regularity of the deformation fields as detailed in Sec. 2.2. An image-adaptive minimum-spanning-tree is used as a graph-model, which enables both a direct (non-iterative) solution and a suitable model for discontinuous sliding motion [3, 6]. Finally, a multi-scale, symmetric transformation model is used to recover both small and large scale motion and avoid singularities in the deformation field [7].

2.1 Image Similarity using Self-Similarity Context

The self-similarity context (SSC) is a multi-dimensional image descriptor, which is estimated based on patch-based self-similarities in a similar way as in our

previous work, the modality independent neighbourhood descriptor (MIND) [4, 5]. But instead of extracting a representation of local shape or geometry, SSC aims to find the context around the voxel of interest. Therefore, the negative influence of noisy patches can be greatly reduced, making this approach very suitable for low-dose CT registration.

Self-similarity can be described by a distance function between image patches within one image I (sum of squared differences SSD can be used within the same scan), a local or global noise estimate var, and a certain neighbourhood layout $\mathcal{N}$ for which self-similarities are calculated. For a patch centred at $\mathbf{x}$ the self-similarity descriptor $\mathcal{S}(I, \mathbf{x}, \mathbf{y})$ is given by:

$$\mathcal{S}(I, \mathbf{x}, \mathbf{y}) = \exp(-\frac{SSD(\mathbf{x}, \mathbf{y})}{var}) \quad \mathbf{x}, \mathbf{y} \in \mathcal{N} \tag{1}$$

where $\mathbf{y}$ defines the centre location of a patch within $\mathcal{N}$. In [5] and [9], the neighbourhood layout was defined to always include the patch centred around $\mathbf{x}$ for pairwise distance calculations. This has the disadvantage that image artefacts or noise within the central patch always have a direct adverse effect on the self-similarity descriptor. We therefore propose to avoid using the central patch and instead use all pairwise distances of patches within the six neighbourhood (with a Euclidean distance of $\sqrt{2}$ between them). The noise estimate var in Eq. 1 is defined to be the mean of all patch distances and descriptors are normalised so that $\max(\mathcal{S}) = 1$.

Once the descriptors are extracted for both images, yielding a vector for each voxel, the similarity metric between locations $\mathbf{x}_i$ and $\mathbf{x}_j$ in two images I and J can be defined as the sum of absolute differences (SAD) between their corresponding descriptors. The distance D between two descriptors is therefore:

$$D(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{|\mathcal{N}|} \sum_{\mathbf{y} \in \mathcal{N}} |\mathcal{S}(I, \mathbf{x}_i, \mathbf{y}) - \mathcal{S}(J, \mathbf{x}_j, \mathbf{y})| \tag{2}$$

Equation 2 requires $|\mathcal{N}|$ computations to evaluate the similarity at one voxel. The discrete optimisation frameworks used in our work, requires many cost function evaluations per voxel. In order to speed up the computations, we quantise the descriptor to a single integer value with 64 bits, without significant loss of accuracy. The exact similarity evaluation of Eq. 2 can then be obtained using the Hamming distance between two descriptors using only one operation per voxel. A descriptor using self-similarity context consists of 12 elements, for which we use 4 bits per element, which translates into 5 different possible values. For the EMPIRE challenge data, we add an additional term, which is directly based on the intensity difference between scans. This is used to improve the alignment of the lung boundary, which is challenging for our approach, because we do not make use of any lung segmentations. Specifically, the intensity values of the scans are quantised into 15 bits within a range of $[-800, -100]$ Hounsfield units and amended to the SSC descriptor.

2.2 Discrete Optimisation and Dense Displacement Sampling

Discrete optimisation is used in this work, because it is computational efficient, no derivative of the similarity cost is needed, local minima are avoided and large deformations can be covered by defining an appropriate range of displacements $\mathbf{u}$. We use the framework presented in [3, 6], where a graph is defined, in which the nodes $p \in \mathcal{P}$ (with spatial location $\mathbf{x}_p$) correspond to control points in a uniform B-spline grid. For each node, there is a set of labels f_p, which correspond to a three dimensional displacement $f_p = \mathbf{u}_p = \{u_p, v_p, w_p\}$ between a control point p in the fixed image I and the moving image J. The energy function to be optimised consists of two terms: the similarity cost D and the pair-wise regularisation cost $R(f_p, f_q)$ for any node q, which is directly connected ($\in \mathcal{N}$) with p:

$$E(f) = \underbrace{\sum_{p\in\mathcal{P}} D(f_p)}_{\text{similarity term}} + \alpha \underbrace{\sum_{(p,q)\in\mathcal{N}} R(f_p, f_q)}_{\text{regularisation term}} \tag{3}$$

The weighting parameter α sets the influence of the regularisation. As discussed before SSC is used as similarity metric and the deformation field is regularised using the squared differences of the displacements of neighbouring control points:

$$R(f_p, f_q) = \sum_{(p,q)\in\mathcal{N}} \frac{||\mathbf{u}_p - \mathbf{u}_q||^2}{||\mathbf{x}_p - \mathbf{x}_q||} \tag{4}$$

This diffusion regularisation is always applied to full current deformations (incremental regularisation, as detailed in [3] Sec. II. C). In [3] we showed that medical images can be efficiently treated using a relaxed graph structure. Instead of connecting each node to its six immediate neighbours, only the most relevant edges are considered, where the edge weight $w(p, q)$ is defined as the sum of absolute differences (SAD) between the intensities of all voxels within the influence regions of two neighbouring control points p and q. By applying Prim's algorithm, we obtain to a minimum spanning tree (MST) with minimum total edge costs. Message passing (belief propagation) on a tree graph (BP-T) [2] finds the global minimum, without iterations, in only two passes.

The challenges of the dataset require us to optimise over a large number of degrees of freedoms. The moving scans (as defined by the dataset) are resampled to a resolution of $0.8\times0.8\times0.8$ mm (or $0.6\times0.6\times0.6$ mm for scans with very high original resolution). The moving scans are resampled to a resolution so that the image dimensions are matched to deal with different dimensions of corresponding scan pairs. A uniform grid of control points with levels of decreasing grid-spacing of $g = \{6, 5, 4, 3, 2\}$ voxels is employed. This means that the number of nodes is increased for each subsequent level. The number of labels is correspondingly decreased, and the maximum search radius is set to $l_{\max} = \{6, 5, 4, 3, 2\}$ steps. The search space is defined as $\mathcal{L} = d \cdot \{0, \pm1, \ldots, \pm l_{\max}\}^3$ voxels, where d is a discretisation step, which is defined as $\{5, 4, 3, 2, 1\}$ voxels for the five levels.

In order to further reduce the computation time of the similarity not all voxels within the influence region of each control point, but only a regularly spaced

subset (≈ 50) of them is taken into account. We have shown in [3] that this subsampling has no adverse effect on the quality of the estimated deformations and results in a 2.5× speed-up of the similarity term calculations.

Inverse consistent mappings are beneficial in order to avoid singularities and remove a potential bias of the selection of moving and fixed image. We estimate both the forward and backward displacement fields $\mathbf{u}^n$ and $\mathbf{v}^n$ independently. Then a simple scheme can be used to ensure inverse consistency by iteratively updating the following equations (here 20 iterations are used):

$$\begin{aligned} \mathbf{u}^{n+1} &= 0.5(\mathbf{u}^n - \mathbf{v}^n(\mathbf{x} + \mathbf{u}^n)) \\ \mathbf{v}^{n+1} &= 0.5(\mathbf{v}^n - \mathbf{u}^n(\mathbf{x} + \mathbf{v}^n)) \end{aligned} \tag{5}$$

An example of a deformable registration of a 4D-CT scan pair (# 23 of the EMPIRE challenge dataset) is shown in Fig. 1.

3 Implementation

Our approach does not require the use of lung segmentations or a linear pre-registration step. It would also naturally extend to the use for multi-modal scans (when removing the additional intensity based similarity term). The computation time for a full registration of a typical scan pair is only three minutes (on a quad-core CPU), which is orders of magnitudes faster than most other algorithms used for the challenge. Approximately a third of the time is spend each for extracting the SSC descriptors, calculating the similarity terms and optimisation of the regularised energy using BP-T. The method is implemented in C++ and OpenCL, which makes use of single instruction multiple data (SIMD) for intra-core parallelism, but could also directly be used on graphics processing units (GPU) to further reduce computation times. The code is freely available at `http://users.ox.ac.uk/~shil3388`, to allow for an exact replication of the presented results.

In addition to the parameter settings mentioned above, we use a regularisation weighting of $\alpha = 2$. This values has been optimised for the lowest registration error for the DIR-lab COPD dataset [1], although a wide range $[0.5, 8]$ of other values gives very similar results. The patch-size for the distance calculations for the SSC descriptor is chosen to be $7 \times 7 \times 7$ voxels for the first level, $5 \times 5 \times 5$ voxels for the second and third levels and $3 \times 3 \times 3$ voxels for the remaining levels. Descriptors for both images are recalculated (when necessary) for each level.

4 Discussion

We have presented a flexible, fast algorithm for deformable intra-patient lung registration. It does not require the use of additional anatomical knowledge (such as lung segmentations), has a run-time of only three minutes (on a quad-core

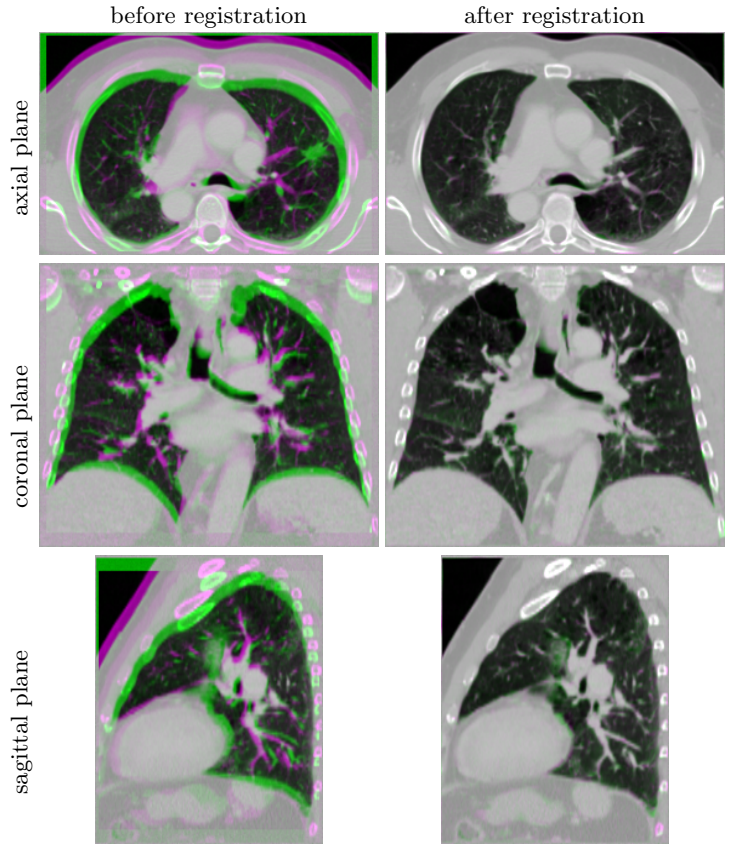

Fig. 1. Deformable registration of 4D-CT scan pair (# 23 of [8]) using SSC and **deeds**. The inhale phase is shown in green with an overlay of the exhale phase in magenta. A visually good alignment of inner-lung structures (such as vessels and airways) as well as the rib cage outside the lungs can be seen.

CPU) and achieves state-of-the-art registration quality. We found that while simple intensity based similarity metrics appear to work well for scans with lower motion amplitude (inhale/inhale, 4DCT), they fail for cases with large displacements (inspiration/expiration). Here, the use of the self-similarity context (SSC), which was originally intended for multi-modal registration, has great benefits, because it is more discriminative for important anatomical features and robust to noise and locally varying contrast. When estimating motion for the whole image domain, without restricting ourselves to the alignment of inner-lung structures (as done by all of the algorithm currently ranked in the top 15), the dense displacement sampling in combination with the approximately globally optimal regularisation enables good alignment of all image features even for cases of large motion. To deal with the particularly challenging sliding motion, a sufficiently fine grid-spacing should be employed. The registration accuracy could be further

improved by using the provided lung masks and/or defining a finer grid-spacing and smaller quantisation steps.

References

1. Castillo, R., Castillo, E., Fuentes, D., Ahmad, M., Wood, A.M., Ludwig, M.S., Guerrero, T.: A reference dataset for deformable image registration spatial accuracy evaluation using the COPDgene study archive. Physics in medicine and biology 58(9), 2861 (2013)
2. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient matching of pictorial structures. In: Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on. vol. 2, pp. 66–73. IEEE (2000)
3. Heinrich, M., Jenkinson, M., Brady, S., Schnabel, J.: MRF-based deformable registration and ventilation estimation of lung CT. Medical Imaging, IEEE Transactions on (2013)
4. Heinrich, M.P., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F.V., Brady, J.M., Schnabel, J.A.: Non-local shape descriptor: a new similarity metric for deformable multi-modal registration. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011, pp. 541–548. Springer (2011)
5. Heinrich, M.P., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F.V., Brady, S.M., Schnabel, J.A.: MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration. Medical Image Analysis (2012)
6. Heinrich, M.P., Jenkinson, M., Brady, M., Schnabel, J.A.: Globally optimal deformable registration on a minimum spanning tree using dense displacement sampling. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012, pp. 115–122. Springer (2012)
7. Heinrich, M., Jenkinson, M., Papiez, B., Brady, M., Schnabel, J.: Towards real-time multi-modal fusion for image-guided interventions using self-similarities. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013. Springer (2013)
8. Murphy, K., Van Ginneken, B., Reinhardt, J.M., Kabus, S., Ding, K., Deng, X., Cao, K., Du, K., Christensen, G.E., Garcia, V., et al.: Evaluation of registration methods on thoracic CT: The EMPIRE10 challenge. Medical Imaging, IEEE Transactions on 30(11), 1901–1920 (2011)
9. Shechtman, E., Irani, M.: Matching local self-similarities across images and videos. In: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on. pp. 1–8. IEEE (2007)