

Supervoxels for Graph Cuts-Based Deformable Image Registration Using Guided Image Filtering

Adam Szmul^{a,*}, Bartłomiej W. Papież^a, Andre Hallack^a, Vicente Grau^a, Julia A. Schnabel^{a,b}

^aInstitute of Biomedical Engineering, Department of Engineering Science, University of Oxford, UK

^bDepartment of Biomedical Engineering, Division of Imaging Sciences and Biomedical Engineering, King's College London, UK

Abstract. In this work we propose to combine a supervoxel-based image representation with the concept of graph cuts as an efficient optimization technique for 3D deformable image registration. Due to the pixels/voxels-wise graph construction, the use of graph cuts in this context has been mainly limited to 2D applications. However, our work overcomes some of the previous limitations by posing the problem on a graph created by adjacent supervoxels, where the number of nodes in the graph is reduced from the number of voxels to the number of supervoxels. We demonstrate how a supervoxel image representation, combined with graph cuts-based optimization can be applied to 3D data. We further show that the application of a relaxed graph representation of the image, followed by guided image filtering over the estimated deformation field, allows us to model ‘sliding motion’. Applying this method to lung image registration, results in highly accurate image registration and anatomically plausible estimations of the deformations. Evaluation of our method on a publicly available Computed Tomography lung image dataset (www.dir-lab.com) leads to the observation that our new approach compares very favorably with state-of-the-art in continuous and discrete image registration methods achieving Target Registration Error of 1.16mm on average per landmark.

Keywords: supervoxels, image registration, graph cuts, guided image filtering, lung motion.

*Adam Szmul, adam.szmul@eng.ox.ac.uk

1 Introduction

Superpixel clustering algorithms perform segmentation of images into anatomically and visually meaningful regions. They have found a wide range of application in imaging, including object localization and class segmentation,¹ human pose estimation² or depth estimation.^{3,4} While the traditional application of superpixels is to image (over)segmentation, here we employ them to remove redundant image intensity information in order to reduce the number of nodes needed for the graph-based registration. At the same time, as superpixels tend to preserve consistent edges, this allows us to effectively preserve sliding boundaries between organs. For 3D application, the clusters represent groups of voxels and therefore are called supervoxels. In this work, we present their application to medical imaging, which is an inherent part of the patient diagnostic, treatment

and follow-up care. Modern methods enable acquisition of many images of the same subject at different time points and with the use of different scanners. To take full advantage of such possibilities, it is necessary to find spatial correspondence between the images, which allows us for an assessment of the same structures in the different acquired images. The process of finding an optimal transformation between different images is called image registration and especially for medical applications, remains an active and still challenging research field.⁵ One of the most demanding applications of image registration is lung image registration, particularly due to changes in lung density during breathing⁶ and to the relative motion⁷ between the lungs and surrounding tissues. These challenges render most of the single-modality similarity measures along with conventional regularizers for deformable image registration unsuited to the task. The relative motion between the lung and surrounding tissues is usually referred to as 'sliding motion' in the literature and is characterized by discontinuities in tangential components of the displacement field along the direction normal to the lung surface, while tangential components in the direction tangential to the surface, as well as normal components, remain smooth. One of the methods addressing this challenge is to perform lung segmentation before the registration and restrict the registration only to the lungs. However, such an approach has the limitation that it requires accurate lung masks and neglects intralobar sliding motion, which might occur inside of the lungs. Among different imaging modalities applicable to lung imaging, Computed Tomography (CT) remains the most widely used volumetric imaging modality. This is the reference modality for the particular case of lung cancer, playing a key role in radiotherapy planning⁸ and treatment monitoring.⁹ In medical imaging, it is difficult for the clinical expert to directly compare image volumes acquired at different times. In order to accurately assess the same structures, a very careful image registration needs to be performed to allow for direct side-by-side comparison. Practical applications for this, in the context of

lung imaging, is image-guided radiotherapy, where the dose plan is devised from a baseline scan, and mapped, using image registration, to the patient for more accurate dose delivery (a concept known as dose painting), taking the different breathing state into account. Another application is monitoring of response to radiotherapy, where in addition to compensating differences in patient breath hold, actual morphological or functional changes need to be assessed. For such sophisticated tasks, voxel size or even sub-voxel image registration accuracy would be expected. It can also be used for lung ventilation quantification,¹⁰ which has the potential to spare well-functioning parts of the lungs during radiotherapy. In the case of the lung ventilation quantification based on 4DCT, an accurate registration is crucial. The most accepted approach tracks changes in lung tissue intensity, which directly correspond to the tissue density. Existing methods perform an initial Gaussian image smoothing to compensate for errors in registration. Having images accurately registered, such smoothing can be avoided, resulting in more detailed ventilation maps. The estimated ventilation maps could be further used to correct the radiotherapy plan to avoid side effects due to irradiation of healthy tissue.

1.1 Deformable Image Registration

In general, in deformable image registration, one of the images is considered to be fixed, I_{fix} , and another moving, I_{mov} , which is transformed towards the fixed one. This process can be shown as an energy minimization procedure of a similarity measure of the images, $Sim(I_{fix}, I_{mov}(\mathbf{u}(\mathbf{x})))$, with $I_{mov}(\mathbf{u}(\mathbf{x}))$ standing for I_{mov} warped toward I_{fix} by the deformation \mathbf{u} at a spatial location \mathbf{x} . This term measures the distance between the images based on the chosen similarity measure. However, ~~relying~~ relying purely on similarity measure, which matches spatial locations based on the chosen measure, might result in a physiologically implausible deformations.¹¹ The transformation for

lung registration should preserve discontinuities, modeling sliding motion by allowing for discontinuities at the lung boundaries and between lobes, while providing smooth deformations in other regions. Therefore, the transformation model might be inherently regularized, whereas, in others, an additional regularization term $Reg(\mathbf{u}(\mathbf{x}))$, which penalizes implausible deformations, is introduced to further constrain the transformation model. The exact definition of the $Reg(\mathbf{u}(\mathbf{x}))$ term depends on the application but usually is defined in terms of derivatives of the deformations. The influence of the regularization term is controlled by a weighting parameter κ . The most common formulation of the energy to be minimized takes the form of:

$$E(\mathbf{u}(\mathbf{x})) = Sim\left(I_{fix}, I_{mov}(\mathbf{u}(\mathbf{x})) + \kappa Reg(\mathbf{u}(\mathbf{x}))\right). \quad (1)$$

There have been many different methods proposed for deformable lung image registration, which in general can be categorized in terms of the applied optimization approach into continuous or discrete optimization-based approaches. Among continuous optimization-based approaches for sliding motion preservation, one of the proposed methods was to decouple diffusion regularization, which penalizes the squared magnitude of the gradients of the deformation field,¹² into normal and tangential directions around the lung boundaries using automatically detected masks.¹³ One of the limitations of the method is that it requires accurate lung segmentation for calculating normal and tangential direction vectors. Another one is that it neglects the sliding motion which might occur inside the lungs. Modeling it with that method would require lung segmentation into individual lobes, which is not a trivial task in general, and is even more challenging in 4DCT due to the reconstruction artifacts. A similar approach was used for Positron Emission Tomography (PET) to CT image registration.¹⁴ A piecewise diffeomorphism, as an extension to Large Deformation

Diffeomorphic Metric Mapping (LDDMM), was proposed in Ref. 15, whereas bilateral filtering, as
 regularization of the deformation field preserving sliding motion, was introduced in Ref. 16. These
 methods use Demons-based optimization schemes,¹⁷ which may not be robust enough to local
 minima. Using a discrete optimization-based approach, a Markov Random Field (MRF)-based
 method has been presented in Ref. 18, where the optimization problem is posed on a minimum
 spanning tree extracted from a graph. The image is parametrized to regular cubes represented by
 nodes in the graph. However, the method does not address the sliding motion problem directly.
 More recently, a symmetric regularized correspondence fields method was introduced in Ref. 19
 for images from patients with Chronic Obstructive Pulmonary Disease (COPD) and a supervoxel-
 based belief propagation algorithm in Ref. 20. For the latter, the optimization has also been posed
 on a minimum spanning tree, this time created by a supervoxel image parametrization. However,
 the tree representation, because of the hierarchical nature, misses many of the connections between
 regions either inside or outside of the lungs. This will break up the regularization components that
 we would expect between those regions. In contrast, a graph like the one proposed here preserves
 all relevant connections.

1.2 MRFs and Graph Cuts

To use a discrete solver for deformable image registration, the problem is usually formulated in
 terms of MRFs. The optimization is posed on an undirected graph that is built by a set of nodes
 P . MRFs create a framework where long-range relations of distant nodes in the graph are modeled
 via a *knock-on effect* by short-range relations of usually just a pair of neighbouring nodes. In such
 a form, the optimization is the process of finding a labelling f for the set P that minimizes the

117 energy $E(f)$, consisting of the sum of two terms:

$$E(f) = E_{data}(f) + \alpha E_{smooth}(f), \quad (2)$$

118 where $E_{data}(f)$ represents the disagreement between the labelling f and the observed data, whereas
119 an α -weighted piecewise smoothness term $E_{smooth}(f)$ forces nodes connected by an edge to have
120 similar labels. Depending on the application, the labels can represent different properties, for
121 instance objects or background in segmentation²¹ or a depth map in stereo-vision.²² Such formula-
122 tion can also be applied to deformable image registration, where the labels to be assigned represent
123 predefined displacement vectors $[u_x, u_y, u_z]$. For this application, only a limited number of dis-
124 placements is considered, thus this formulation requires a transformation model, which transfers
125 the displacement field from a discrete domain into a continuous one. However, for a 3D application
126 the number of labels can consist of thousands of allowed displacements.

127 Methods capable of solving such a formulation have attracted researchers since its introduc-
128 tion. However the lack of efficient optimization algorithms limited their application for deformable
129 image registration. An approach using dynamic programming as an optimization method as an op-
130 timization method, **drop**,²³ has been successfully applied for brain image registration. The main
131 limitation of this method is that it estimates the best displacements in the x, y and z directions
132 independently. What is more, it does it in an iterative way, which renders the method similar to
133 the continuous-optimization based approaches and therefore potentially vulnerable to be trapped
134 in a local minimum. These restrictions limits the advantage of the discrete optimization-based
135 approach, which allows for a dense displacement sampling and for some methods to find at least
136 a strong local minimum. In Ref. 24, the authors proposed to use a random walks algorithm for

deformable image registration. Both methods proposed in Refs. 18 and 20 use belief propagation algorithm as the optimization method to estimate displacement field.

Another example of the optimization method are graph cuts. Although the method has been widely used in many fields of computer vision, such as image segmentation,²¹ de-noising,²⁵ stereo vision²⁶ or motion detection,²⁷ its application to image registration has been limited. A voxel-wise deformable registration method applying graph cuts has been proposed in Ref. 28 for both 2D and 3D brain images. An obvious limitation of the method is the size of the graph on which the problem is posed. While for 2D images the problem is still tractable, in 3D cases the size of the graph with every voxel represented by a node. This, along with the high number of possible displacements, make the problem considerably challenging in terms of computation. For instance, the method presented in Ref. 29, which is an extension of the approach from Ref. 28, reports computational time for 3D cases of up to 24h. A landmark-based graph cuts algorithm was proposed for medical image registration in Ref. 30. In that work, a penalty cost is added to the data term for a node on which a landmark is localized, forcing its alignment with the corresponding landmark in the other image. Such an approach can be seen as putting a hard constraint on the deformation field in presence of landmarks. A registration method using mutual information as a similarity measure with graph cuts as an optimization method was later proposed in Ref. 31 and followed by a similar approach with prior joint intensity distribution in Ref. 32.

To address the computational cost issue inherent to pixel/voxel-wise formulations of the problem, here we propose the use of supervoxel image representation.³³ This approach allows us to reduce the complexity of the problem by representing a group of voxels by a single supervoxel. The graph cuts algorithm has a worst-case computational complexity of $O(n^2)$ for a graph with n nodes.³⁴ In our case, the number of nodes n is reduced from the number of voxels to the number of

160 supervoxels. When each of the supervoxels consists of n_s voxels, then the number of supervoxels
 161 $N = \frac{n}{n_s}$, and the worst-case computational complexity of our formulation is expressed as $O(N^2)$.
 162 Subsequently, the number of displacements to be estimated is reduced from the number of vox-
 163 els in the image to the number of extracted supervoxels, which is significantly lower. The use of
 164 supervoxels has certain limitations. Small supervoxels diminish the advantage of the reduction of
 165 the number of nodes in the graph (in the limit, the number of supervoxels can become the number
 166 of voxels). On the other hand, when the supervoxels are large, they might not be able to accu-
 167 rately model the variations in displacement. In the process of finding the optimum displacements,
 168 we apply the graph cuts method and later use an edge-preserving filtering method, namely guided
 169 image filtering,³⁵ to convert the displacement field derived from the sparse image representation
 170 into a piecewise continuous displacement field. The proposed method has been initially presented
 171 in Ref. 36. The paper has been deeply extended and rewritten. In particular, here we present the
 172 method in full detail. Secondly, we provide a comprehensive evaluation of different variants of the
 173 method, justifying the chosen approach. Thirdly, we compare the method against state-of-the-art
 174 including, a regular grid-like image representation and investigate the influence of filtering meth-
 175 ods. Finally, we also show how the method and its variants handle the sliding motion modeling.
 176 The closest method in the field to the proposed is the one presented in Ref. 20. Even though both
 177 methods share some common features, such as the supervoxel approach to the image parametriza-
 178 tion or discrete optimization, closer analysis reveals that the methods present completely different
 179 approaches to image registration. In our method we use a multi-resolution approach to extract su-
 180 pervoxels, whereas in Ref. 20 the supervoxels are extracted from the original, unprocessed images.
 181 In our method, at every resolution level we extract supervoxels with different relative size, starting
 182 with larger ones and gradually reducing their size in following resolution levels. Another major

difference lies in the graph representation, which will be presented in Section 2, along with the chosen method to convert the sparse displacement field into the locally continuous one.

1.3 Aims

The main aims of this paper are fourfold:

1. To propose a method for reduction of the size of the 3D deformable image registration problem, making it efficiently solvable via the graph cuts method by using supervoxels.
2. To introduce a novel graph relaxation criterion based on a relation between intensities of adjacent supervoxels, which improves the 3D deformable image registration performance for the lungs application. We also show that it is capable of preserving well the sliding motion of the lungs.
3. To show that the proposed framework is robust to the supervoxels extraction and the graph relaxation parameters, providing reasonably good results , in terms of registration accuracy, for a range of parameters and modifications to the graph connectivities, as well as applied filtering methods.
4. To demonstrate that the proposed framework performs favorably to the most similar deformable image registration methods, especially well to those which do not use supervoxel image representation, when compared on a publicly available data set.

The remainder of this paper is structured as follows. In Section 2 we describe our method, introducing all of its components. In Section 3 we present the conducted experiments and obtained results. The discussion about the results, plans for future work and conclusions are presented in Section 4.

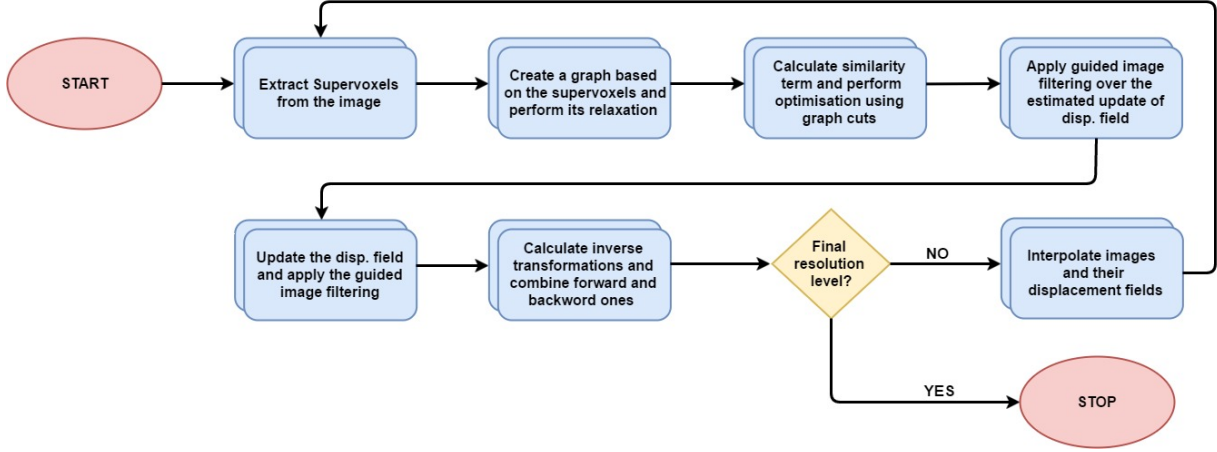


Fig 1 Diagram presenting the workflow of the proposed method. We start from extracting supervoxels and creating a graph based on them. Then we estimate a displacement for each supervoxel using the graph cuts optimization method. The guided image filtering is first applied to the estimated update of the displacement field, similar to fluid-like regularization in Demons, and then to the update combined with the previously estimated displacement field, similar to diffusion-like regularization in Demons. We then estimate inverse transformations and combine them with forward ones. Double rectangles emphasize the symmetric nature of the method, where forward and backward transformation from the target to the moving image are estimated at the same time.

2 Methods

In this section, we present the proposed method for deformable image registration, using supervoxel-based image representation and graph cuts as an optimisation. An overview of the proposed method is shown in Figure 1, with every component addressed individually in the following sections. In Section 2.1, we explain the way in which the images are over-segmented into supervoxels and describe the selected image clustering method. This is followed by the introduction of the similarity measure of choice in Section 2.2. A detailed explanation of the graph construction and energy function formulation used for the initial displacement field estimation is presented in Section 2.3. Section 2.4 describes the proposed transition from the discrete domain to the continuous with use of the guided image filtering method. In the Section 2.5 we explain how we combine the estimated forward and backward transformations.

2.1 Image Clustering

The simplest way to reduce the complexity of deformable image registration, due to the size of the images, is to parameterize it using a regular grid structure. Even though such an approach efficiently reduces the complexity of the problem, it does not capture well the semantic regions of the image, usually represented by higher contrast edges. To overcome this limitation, superpixel algorithms were proposed, which group pixels with similar appearance within a local neighbourhood. In the literature, many approaches have been proposed to perform image clustering, including mean-shift methods,²² gradient-based³⁷ and the watersheds method.³⁸ However, probably the most popular image clustering method is Simple Linear Iterative Clustering (SLIC),³³ mainly due to the speed of performance, direct control over the number of extracted supervoxels and supervoxel compactness. The SLIC method is designed to extract k approximately equally-sized supervoxels. It starts by distributing seeds for each of the supervoxels, initially placed at intervals $S = \sqrt[3]{M/k}$ voxels apart, with M being the total number of voxels in the image. The positions of the centers are then corrected based on the gradients of the image to avoid locating them on image edges or at a noisy voxel before each voxel is assigned to the nearest cluster. In an iterative manner, the distance between each voxel \mathbf{x} and the closest cluster center \mathbf{c} is calculated based on the Euclidean distance $d_e = \|\mathbf{x} - \mathbf{c}\|$ and the intensity-based similarity is $d_I = \|(I(\mathbf{x}) - I(\mathbf{c}))\|$. The distance \mathcal{V} is defined as $\mathcal{V} = ((d_e)^2 + (\frac{d_I}{S})^2 m^2)^{1/2}$ where m is a parameter, which controls the compactness of supervoxels.

2.2 Similarity Measure

The simplest similarity measures, ~~relying~~ **relying** purely on intensities are not well-suited for CT lung registration, mainly due to changes in lung density during breathing, resulting in a variation

in their intensity. They are also inadequate for multi-modal image registration, where intensities represent different physical properties. In our method we propose to use the Modality Independent Neighbourhood Descriptor (MIND),³⁹ which has been originally developed for lung image registration. The descriptor is defined for each voxel position \mathbf{x} with a spatial search region R by a relation between a distance D_p and a variance Var :

$$\text{MIND}(I, \mathbf{x}) = \frac{1}{\eta} \exp \left(-\frac{D_p(I, \mathbf{x}, \mathbf{x} + \mathbf{r})}{Var(I, \mathbf{x})} \right) \quad (3)$$

with η normalizing the values to 1. The distance D_p measures the intensity difference between the voxels in a defined range \mathbf{r} :

$$D_p(I, \mathbf{x}, \mathbf{x} + \mathbf{r}) = G * (I(\mathbf{x}) - I(\mathbf{x} + \mathbf{r}))^2 \quad (4)$$

with G being a Gaussian kernel. The variance $Var(I, \mathbf{x})$ in the denominator stands for the estimation of intensity distribution in the neighbourhood of a voxel \mathbf{x} and is represented as the variance in intensities within a neighbourhood N . For more details of the method we refer the interested to Ref. 39. The main advantage of the MIND over single value intensity-based similarity measures is that it provides a descriptor, which is sensitive to similarities in gradients and texture but insensitive to direct differences in intensities between voxels. These properties are extremely useful for lung image registration, where the lungs change their density, and therefore their appearance in the CT images, during the breathing process.

2.3 Energy Formulation

Our optimization problem is posed on an undirected graph, defined by the adjacency of previously extracted supervoxels. In the graph, each supervoxel is represented by a node and every pair of directly adjacent supervoxels is connected by an edge. The edges indicate a relation between the supervoxels while their values represent the strength of the relation. The values of the edges are calculated based on the absolute difference between the mean intensities of the supervoxels. Such construction of the graph favors smooth changes between supervoxels having similar mean intensity values and at the same time representing the same structures. To allow for discontinuities in the deformation field between supervoxels clustering different organs, while at the same time preserving the smooth deformation field inside them, we propose to use a relaxed form of the graph. The edges with lower values are assumed to connect different structures, and can thus be removed, based on a relaxation threshold parameter v :

$$V(p, q) = \begin{cases} 1 - \|\overline{I_{mov}(\mathbf{x}_p)} - \overline{I_{mov}(\mathbf{x}_q)}\| & \text{for } \|\overline{I_{mov}(\mathbf{x}_p)} - \overline{I_{mov}(\mathbf{x}_q)}\| < v \\ 0 & \text{for } \|\overline{I_{mov}(\mathbf{x}_p)} - \overline{I_{mov}(\mathbf{x}_q)}\| \geq v, \end{cases} \quad (5)$$

where $\overline{I_{mov}(\mathbf{x}_p)}$ and $\overline{I_{mov}(\mathbf{x}_q)}$ stand for mean intensity values of all voxels $\mathbf{x}_p, \mathbf{x}_q$ contained within the supervoxels represented by nodes p and q , respectively, calculated for the moving image. The image intensities are initially normalised between 0 and 1, therefore the relaxation threshold parameter v is a constant and lies within these ranges. The relaxed graph representation has an additional advantage, as there are fewer non-zero edges remaining in the graph, the overall complexity of the graph layout is reduced and therefore the optimization process is performed with a better rate of the convergence. An example of such clustering and a graph created over the image,

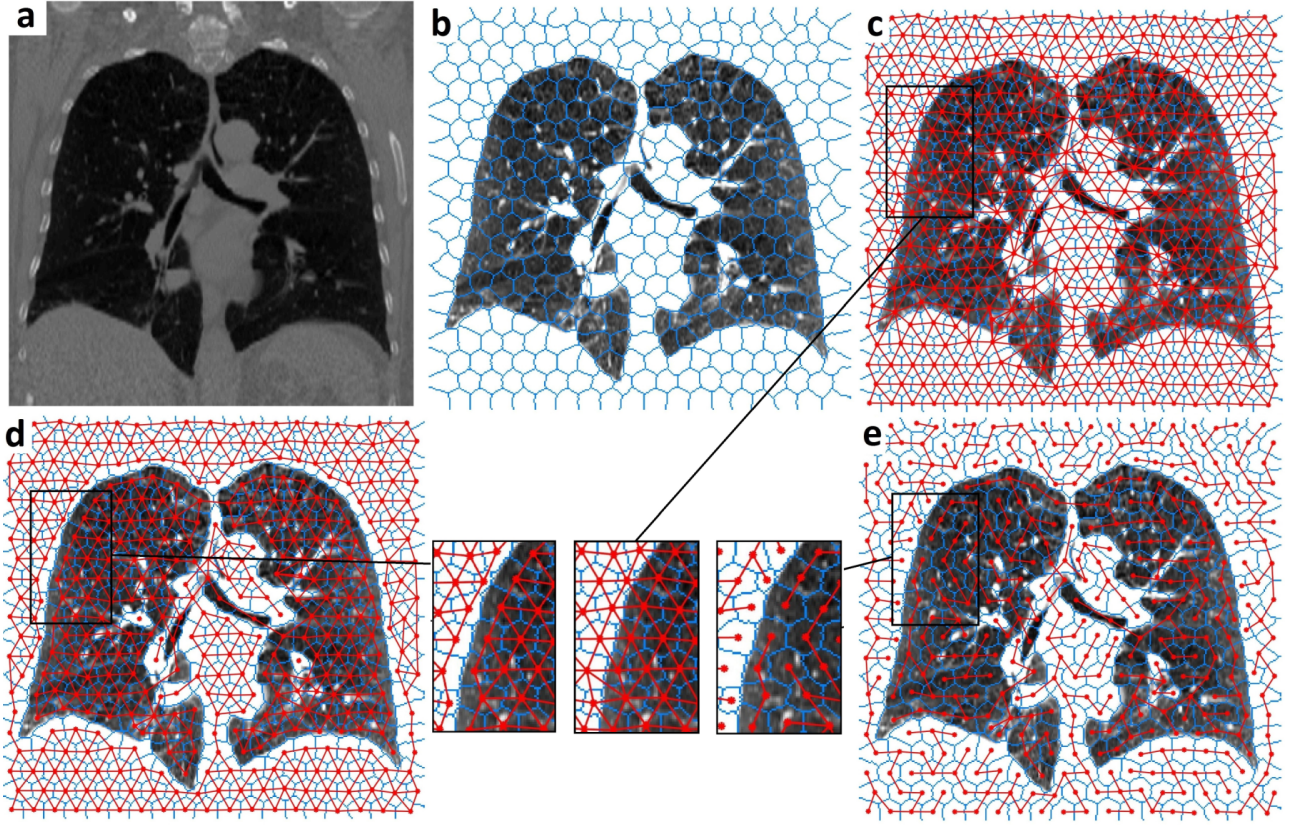


Fig 2 Example of the SLIC algorithm clustering of the peak inhale 2D lung CT image. In (a) the original CT image is shown, (b) shows the corresponding supervoxel representation, with the borders of the supervoxels drawn in blue. The graph extracted based on the clustering is shown in figure (c), with the edges connecting adjacent supervoxels in red. The relaxed graph representation is shown in figure (d), whereas the minimum spanning tree-based simplification of the graph is shown in figure (e). In all of the images, apart from the first one (a), the contrast has been adjusted for better visualisation of supervoxels and graphs. For illustrative purposes, we use 2D superpixels and present them on a 2D slice. However, in the framework we extract supervoxels from 3D volumes.

as well as its relaxed form is shown in Figure 2. The proposed formulation should result in physically plausible deformation inside the lungs and, in turn, a more realistic deformation field. The introduced relaxation of the graph can be considered as a form of pre-segmentation of the image, akin to lung masking, which is a common practice in the field of lung registration.^{15,40,41}

The data cost term is formulated as a mean error calculated for all voxels \mathbf{x} in the fixed image I_{fix} and moving image I_{mov} clustered in a certain supervoxel represented by a node p , for the applied displacement f_p :

$$E_{data}(f) = \sum_p |\overline{\text{MIND}(I_{fix}, \mathbf{x}_p) - \text{MIND}(I_{mov}, \mathbf{x}_p + f_p)}|. \quad (6)$$

We are not matching supervoxels between two images but are finding the best displacement of each of the supervoxels based on the mean error of all their included voxels (3D image pixels) for each of the label, representing a displacement. Such formulation is more robust to the differences in shape of supervoxels across the two images.

Many forms of the piecewise smoothness term for MRFs have been proposed, starting from the simplest Potts model, where the term takes the value 1 when labels are equal and 0 otherwise. In a linear model, the smoothness term changes linearly according to the distance between labels. These formulations of the piecewise smoothness term satisfy all metric requirements, including the triangle inequality. The most common way of formulating regularization in deformable image registration is a quadratic model, which depends on a squared distance between labels f_p and f_q assigned to two neighbouring nodes p and q , defined as:

$$\|f_p - f_q\|^2 = (u_{x_p} - u_{x_q})^2 + (u_{y_p} - u_{y_q})^2 + (u_{z_p} - u_{z_q})^2. \quad (7)$$

289 To recall, f_p is a label representing a displacement vector (x_p, y_p, z_p) assigned to the node p . In our
 290 formulation, we propose the use of the following piecewise smoothness term:

$$E_{smooth}(f) = \sum_{p,q \in N} (1 - \|I_{mov}(\mathbf{x}_p) - I_{mov}(\mathbf{x}_q)\|) \|f_p - f_q\|^2, \quad (8)$$

291 where the $(1 - \|I_{mov}(\mathbf{x}_p) - I_{mov}(\mathbf{x}_q)\|)$ stands for the edge values of the graph and $\|f_p - f_q\|^2$
 292 represents quadratic distance between labels.

293 On such formulated problem we apply the graph cuts method.⁴² Due to the fact that we are
 294 using a quadratic regularization term, which does not satisfy the triangle inequality condition for
 295 metrics, we use the $\alpha - \beta$ swap variant of the graph cuts method,⁴² which has relaxed require-
 296 ments about piecewise smoothness term properties and can be used with a semi-metric piecewise
 297 smoothness term.

298 In our formulation we use a multi-resolution approach. At each resolution level we extract
 299 a number of supervoxels containing a specified number of voxels. For each of the supervoxels
 300 we calculate an error vector with length equal to number of labels. Each value of the vector corre-
 301 sponds to the data term of the energy for a particular displacement represented by the label. In such
 302 a way, we create a data cost matrix, whose size is $(number\ of\ supervoxels) \times (number\ of\ labels)$.
 303 We can see that the optimisation complexity has been heavily reduced compared to the voxel-wise
 304 formulation, where the data cost matrix would have size $(number\ of\ voxels) \times (number\ of\ labels)$,
 305 bearing in mind that every supervoxel represents a group of voxels.

2.4 Guided Image Filtering

The labelling procedure, described in the previous section, results in all supervoxels having assigned a displacement vector, which estimates their best displacement and creates a displacement field. A direct application of the deformation estimated for a supervoxel to all of its voxels might result in foldings at borders of supervoxels that have different deformations, and thus physiologically implausible deformations. To overcome this limitation, in Ref. 20 the authors proposed using multiple layers of supervoxels while in Refs. 23 and 18 a free-form deformation (FFD) with B-splines⁴³ is used as an interpolation method. The properties of B-splines, however, do not allow us directly to model discontinuities of the deformation field, which occur between the lungs, rib cage and diaphragm.

In this work we propose an alternative solution: to smooth the displacement field, while at the same time preserving some of the discontinuities. Guided image filtering image filtering,³⁵ where an output image I_{out} is a linear combination of an input image I_{in} and an image I_g used for guidance:

$$I_{out}(\mathbf{x}) = \sum_{\mathbf{y} \in \mathcal{N}(\mathbf{x})} \mathbf{W}_{\mathbf{y}}(I_g) I_{in}(\mathbf{y}), \quad (9)$$

where $\mathbf{W}_{\mathbf{y}}$ is a filter kernel calculated for the guide image I_g in a neighbourhood $\mathcal{N}(\mathbf{x})$ of a voxel \mathbf{x} . The kernel $\mathbf{W}_{\mathbf{y}}$ is defined as:

$$\mathbf{W}_{\mathbf{y}}(I_g) = 1 + (I_g - \mu_{I_g})^T (I_g - \mu_{I_g}) (\text{cov}_{I_g} + \sigma \mathbf{I})^{-1}, \quad (10)$$

where μ_{I_g} , cov_{I_g} are the mean and covariance of the guidance image I_g calculated in a neighbourhood $\mathcal{N}(\mathbf{x})$, \mathbf{I} is the identity matrix and σ is the smoothness parameter. In our method, this

filtering method is applied over the dense displacement field \mathbf{u} derived from the sparse representation using the moving image I_{mov} for guidance. The discontinuities in the intensity, especially at the borders of organs, such as the lungs, might result in discontinuities in the displacement field. On the other hand, small changes in the intensities between adjacent supervoxels might indicate that they are part of the same organ. This yields smooth deformations across anatomically consistent regions in the moving image, while at the same time preserves discontinuities in the deformation field at the region boundaries. The idea of guided image filtering is based on the edge-preserving approach, similar to bilateral filtering,⁴⁴ which discourages smoothing across boundaries and has good edge-preserving smoothing properties. The main benefit of the method is that it is several times faster than bilateral filtering. Related approaches that were previously proposed include continuous optimization-based registration for the lungs,¹⁶ where bilateral filtering was applied, and liver motion compensation⁴⁵ with guided image filtering.

2.5 Symmetric Registration

In order to enhance the registration process as well as more accurately model the complex nature of the lungs during the breathing process, we propose to apply symmetric registration to our formulation. The registration is performed at different resolution levels and at every level supervoxels of different size were extracted, which was driven by the coarse to fine approach. Therefore, two deformation fields are estimated corresponding to the registrations between moving image towards the target and *vice versa*. Such formulation not only compensates for some mis-registrations originating from the relaxed graph formulation, especially at the border of the lungs and diaphragm, where there is no direct link between the organs and large motion might be observed. It also removes the bias of the choice of the target and source. After performing an

optimization at each resolution level, inverses of both previously estimated deformation fields are approximated based on a simple fixed-point approach originally proposed in Ref. 46 and subsequently combined together:

$$\mathbf{u}_{fix \rightarrow mov}^{new} = 0.5 \cdot (\mathbf{u}_{fix \rightarrow mov} + \mathbf{u}_{mov \rightarrow fix}^{-1}), \quad (11)$$

$$\mathbf{u}_{mov \rightarrow fix}^{new} = 0.5 \cdot (\mathbf{u}_{mov \rightarrow fix} + \mathbf{u}_{fix \rightarrow mov}^{-1}), \quad (12)$$

where $\mathbf{u}_{mov \rightarrow fix}$ and $\mathbf{u}_{fix \rightarrow mov}$ stand for the estimated deformation field transforming I_{mov} to I_{fix} and from I_{fix} to I_{mov} respectively.

3 Experiments and Results

In this section, we evaluate the proposed method on a publicly available CT dataset. We prepared a number of experiments to support the proposed configuration of the method as well as the parameter settings. In Section 3.1 we describe the dataset, which was used for the evaluation purpose. The parameter setting is described in following Section 3.2. In Section 3.3 we explore the influence of the image representation by performing experiments on regular grid image parameterization. This is continued in Section 3.4 with the analysis of the effect on the results of the applied filtering method, by comparing results of Gaussian filtering and the proposed guided image filtering. In Section 3.5 we investigate the influence of the relaxation parameter and in Section 3.6 we show how the method deals with sliding motions. Finally, in Section 3.7 we compare our method against other, most closely related in the field.

3.1 Materials

We evaluated the performance of the proposed image registration method on the publicly available Dir-Lab dataset (www.dir-lab.com),⁴⁷ which consists of 10 cases of dynamic 3D CT volumes covering the full breathing cycle (also called 4D CT) of patients suffering from esophageal or lung cancer. Patient identifiers were removed in accordance with an institutional reviewboard approved retrospective study protocol (RCR 03-0800).⁴⁷ The spatial resolution of the dataset varies between $0.97 \times 0.97 \times 2.5 \text{ mm}^3$ and $1.16 \times 1.16 \times 2.5 \text{ mm}^3$. The dataset provides 300 landmarks for full inhale and full exhale breathing phases, manually placed by an expert. Due to the well-distributed landmarks, this dataset creates a valuable validation framework for deformable lung image registration. To evaluate the performance of the method in terms of registration accuracy, we performed registrations between peak inhale and peak exhale breathing states for the dataset. We calculated the target registration error (TRE) between the annotated landmarks, which is defined as the Euclidean distance over corresponding landmark locations before and after registration. The observer error for this dataset ranges from 0.75mm to 1.13mm.⁴⁷ Such an approach allows us to assess how the methods deals with tracking movement of lung tissue, on which the landmarks were placed.

3.2 Parameter Setting

For all of the experiments we used the same parameter settings for consistency, any changes in the parameter settings are clearly stated. We performed the experiments at 5 resolution levels, at each of them extracting supervoxels of approximate sizes of $[50, 150, 400, 800, 500]$ voxels with maximum allowed displacement of $[3, 4.5, 4.5, 6, 4.5]$ voxels and quantification step of $[0.5, 0.75, 0.75, 1, 0.75]$ respectively. In our implementation of the method, we initially interpolated images to isotropic voxel size for the first four resolution levels. For the final resolution

level, which performs fine-tuning, the method was applied on the original, unprocessed images. The windows w for guided image filtering were set to $[5\ 5\ 5]$, apart from the final resolution level, which was $[7\ 7\ 7]$ voxels and the smoothness parameter σ was set to 0.001. For calculating the MIND descriptor, we used six neighbouring voxels, with range $r=1$. The parameters have been chosen heuristically, based on results on case 8, as the most challenging case in the dataset with the highest initial misalignment. Such an approach should be more robust to over fitting to the data.

3.3 Influence of Image Representation

We conducted an experiment using a regular grid representation, setting SLIC's compactness parameter m to 4000 (in all the other experiments we used $m = 40$), achieving TRE of 1.36mm, compared to 1.16mm for the proposed method. All results are presented in Table 1. Among all of the individual cases the TRE for the proposed method is the lowest. A higher difference was calculated in case 8, where the regular grid variant results in 2.42mm of TRE, whereas for the proposed method TRE was only 1.37mm. In Figure 3 we present the displacement fields estimated by both methods for case 8.

One of the ways to measure the quality of the estimated deformation field might be through calculating the determinant of the Jacobian of the deformations, which can be seen as an indicator of complexity of the deformations. The Jacobian of the deformations is a matrix created by the first derivatives of the deformations calculated for each variable. Its determinant takes the form of:

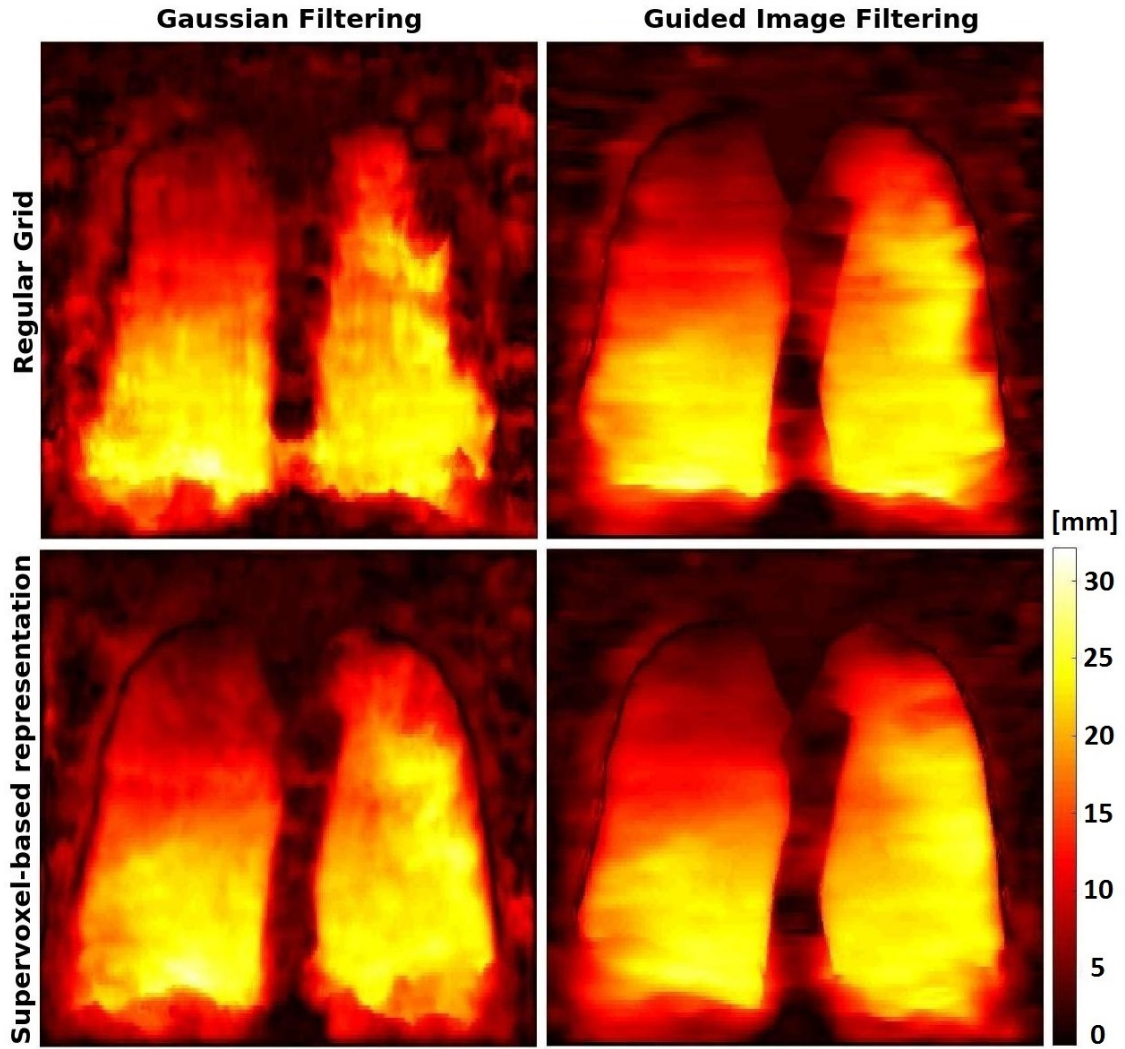


Fig 3 Magnitude of the displacement field in the coronal view for different image representations and filtering methods. On the upper row we show the results for regular grid image representation, while on the lower row for supervoxel-based image representation. The left column presents the results for Gaussian filtering and the right for guided image filtering. The best results were achieved with the proposed configuration shown in the lower right corner.

404

$$\det(Jac(x, y, z)) = \begin{vmatrix} 1 + \frac{\partial u_x(x, y, z)}{\partial x} & \frac{\partial u_x(x, y, z)}{\partial y} & \frac{\partial u_x(x, y, z)}{\partial z} \\ \frac{\partial u_y(x, y, z)}{\partial x} & 1 + \frac{\partial u_y(x, y, z)}{\partial y} & \frac{\partial u_y(x, y, z)}{\partial z} \\ \frac{\partial u_z(x, y, z)}{\partial x} & \frac{\partial u_z(x, y, z)}{\partial y} & 1 + \frac{\partial u_z(x, y, z)}{\partial z} \end{vmatrix}, \quad (13)$$

405 where $u_x(x, y, z)$, $u_y(x, y, z)$, $u_z(x, y, z)$ are the displacement fields estimated in x, y, z directions
406 respectively. In Table 2 we present the fraction of the negative determinant of the Jacobian of
407 deformations, which shows the fraction of voxels for which a one-to-one mapping is violated ⁴⁸,
408 the mean of the determinant of the Jacobian of deformations, showing the change in volume, and
409 the standard deviation of the determinant of the Jacobian of deformations, ~~which can be seen~~
410 ~~as a measure of complexity of the deformations~~, all calculated inside the lungs for the Dir-Lab
411 dataset. We can observe a trend that the variants using supervoxel image representation result in a
412 lower fraction of negative determinant of the Jacobian of deformations, as well as lower standard
413 deviation of the determinant of Jacobian of deformations.

Table 1 Comparison of the proposed method against regular grid image representation, based on TRE in [mm] and standard deviation for the Dir-Lab dataset. The best results are shown in bold that are consistently best for the proposed method using supervoxels and guided image filtering

Target Registration Error in [mm] and standard deviation				
Case	RG + Gauss	Regular grid (RG)	Proposed + Gauss	Proposed
c1	1.16 ± 0.7	0.98 ± 0.5	1.02 ± 0.5	0.95 ± 0.5
c2	1.16 ± 0.7	1.03 ± 0.7	1.03 ± 0.6	0.94 ± 0.5
c3	1.42 ± 1.0	1.16 ± 0.6	1.17 ± 0.6	1.10 ± 0.6
c4	1.62 ± 1.1	1.37 ± 0.9	1.42 ± 1.0	1.35 ± 0.9
c5	1.70 ± 1.8	1.38 ± 1.3	1.46 ± 1.4	1.33 ± 1.2
c6	1.63 ± 1.5	1.20 ± 0.7	1.25 ± 1.1	1.14 ± 0.7
c7	2.39 ± 3.3	1.58 ± 2.0	1.45 ± 1.6	1.13 ± 0.7
c8	4.54 ± 6.4	2.42 ± 3.9	2.37 ± 3.7	1.37 ± 1.9
c9	1.53 ± 1.2	1.27 ± 0.8	1.21 ± 0.7	1.14 ± 0.6
c10	1.52 ± 1.6	1.18 ± 1.0	1.25 ± 0.9	1.16 ± 0.8
Mean	1.87 ± 1.9	1.36 ± 1.2	1.36 ± 1.2	1.16 ± 0.8

Table 2 Comparison of the proposed method against its variants based on fraction of negative determinant of the Jacobian of deformations, which measures the fraction of voxels for which a one-to-one mapping is violated, mean of the determinant of the Jacobian of deformation, showing the change in volume and standard deviation of the determinant of the Jacobian of deformations, which can be seen as a measure of complexity of the deformations, for the Dir-Lab dataset.

Method	RG + Gauss	Regular grid (RG)	Proposed + Gauss	Proposed
Fraction of neg J.	8.5E-3	1.0E-3	1.8E-3	7.2E-4
Mean Jacobian	0.910	0.918	0.915	0.918
Std of Jacobian	0.329	0.191	0.215	0.162

3.4 Impact of Filtering Method

To show the influence of the filtering method on the results, we conducted experiments using both: Gaussian filtering and the guided image filtering. When supervoxel-based image representation was used, for Gaussian filtering achieved TRE was 1.36mm, whereas for the proposed was 1.16mm. Similar trend has been observed, when we applied both filtering method for the regular grid image representation, resulting with TRE of 1.87mm and 1.36mm for Gaussian and guided image filtering respectively. The results are summarized in Table 1. The visualization of the influence of the filtering method is shown in Figure 3, where we present the proposed method with guided image filtering, the proposed method with Gaussian filtering and regular grid with both filtering methods. The results presented in Table 2 show that guided image filtering provides lower fraction of negative determinant of the Jacobian of deformations as well as lower standard deviation of the determinant of the Jacobian of deformations, regardless of image representation. The best overall results are achieved for our proposed method.

3.5 Graph Relaxation

We performed experiments on the full graph, created by all adjacent supervoxels and after performing graph relaxation with threshold parameter v set to 0.1, by removing edges according to

Table 3 Comparison of the proposed method applied over the full graph without the relaxation criterion, after applying the proposed graph relaxation and after relaxing the graph to a minimum spanning tree (MST), based on TRE and standard deviation for Dir-Lab dataset. Best results are shown in bold.

Target Registration Error in [mm] and standard deviation				
Case	Initial	Full graph	Relaxed graph	MST
c1	3.89 ± 2.8	0.95 ± 0.5	0.95 ± 0.5	0.95 ± 0.5
c2	4.34 ± 3.9	0.94 ± 0.5	0.94 ± 0.5	0.97 ± 0.6
c3	6.94 ± 4.0	1.10 ± 0.6	1.10 ± 0.6	1.09 ± 0.6
c4	9.83 ± 4.8	1.37 ± 0.9	1.35 ± 0.9	1.37 ± 0.9
c5	7.48 ± 5.5	1.35 ± 1.2	1.30 ± 1.2	1.37 ± 1.3
c6	10.9 ± 6.9	1.16 ± 0.7	1.14 ± 0.7	1.21 ± 0.8
c7	11.0 ± 7.4	1.20 ± 0.9	1.14 ± 0.7	1.26 ± 1.0
c8	15.0 ± 9.0	1.58 ± 2.4	1.37 ± 1.7	1.75 ± 2.7
c9	7.92 ± 3.9	1.15 ± 0.6	1.14 ± 0.6	1.17 ± 0.7
c10	7.3 ± 6.3	1.12 ± 0.8	1.16 ± 0.8	1.16 ± 0.8
Mean	8.46 ± 5.4	1.19 ± 0.7	1.16 ± 0.8	1.23 ± 1.0

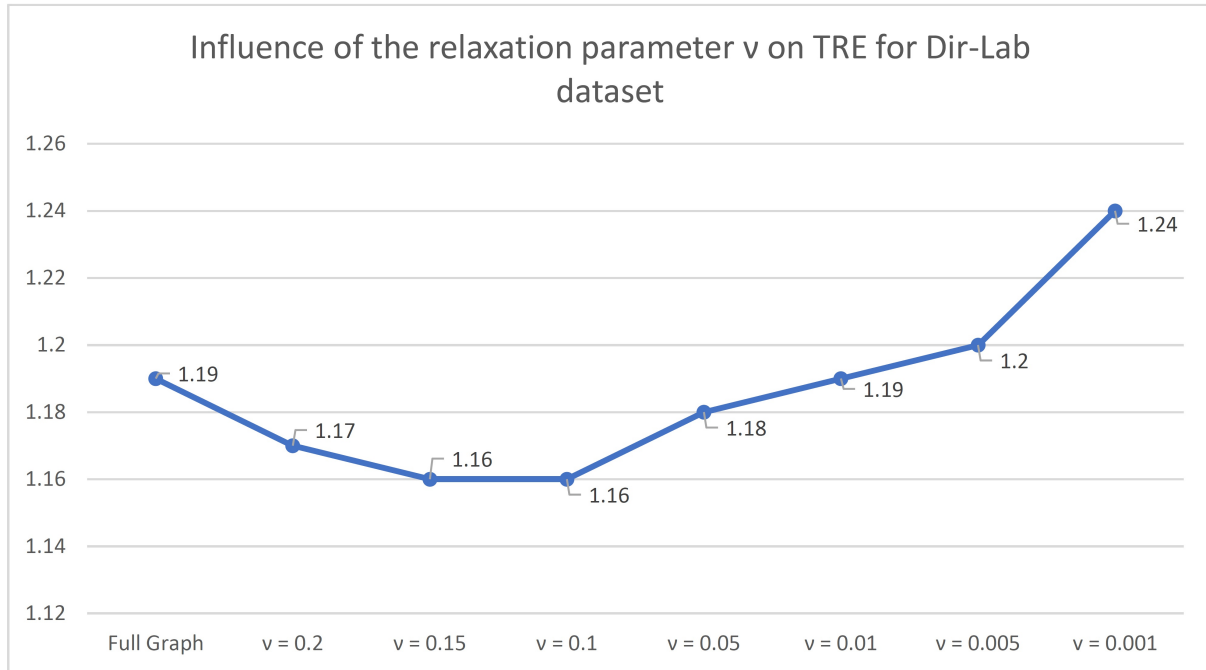


Fig 4 Influence of the relaxation parameter v on TRE in [mm] and standard deviation for the DIR-Lab dataset. The scale in x axis is non-linear

Equation 5. The results for the three first cases were identical in terms of TREs. However, for the remaining six cases the TRE for the relaxed graph representation are lower, apart from the last case 10, where the full graph variant outperformed its relaxed counterpart. We also extracted a MST

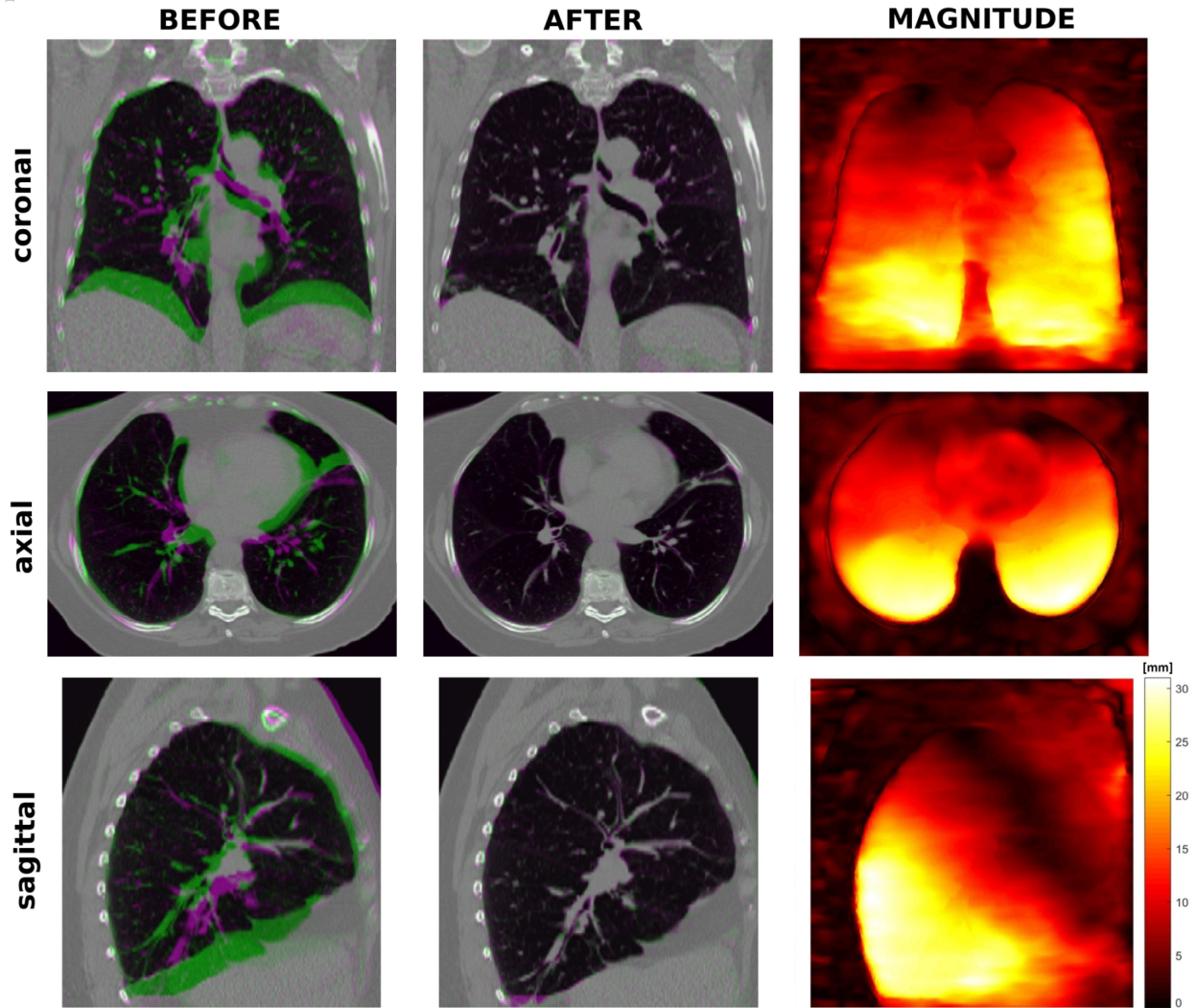


Fig 5 Registration results for the proposed method for the most challenging Dir-Lab case 8 in axial, coronal and sagittal view. In the left column an overlay of full inhale (green) and full exhale (magenta) before applying the registration is presented. In the middle column the same slices after performing the proposed registration are shown. The initial misalignment, visible in the left column, is almost completely eliminated by the proposed registration method. In the right column the magnitude of the displacement field in mm is presented.

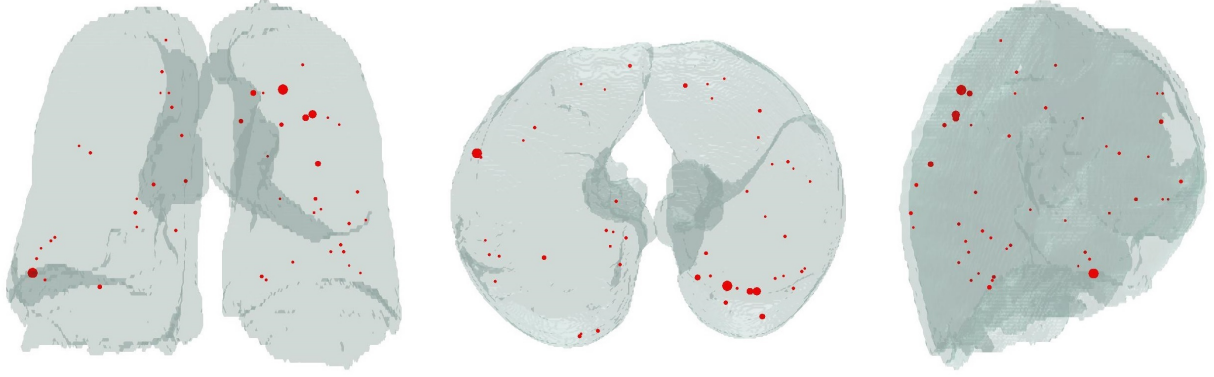


Fig 6 Visualization of the landmark error distribution for the most challenging case 8 after registration using our method in all three views with the lung mask in the background. Only landmarks with an error higher than 1.5mm are displayed. The size of the points corresponds to the landmark error.

from the full graph and performed the optimization using the same energy formulation and parameter settings as for the full graph and its relaxed version and compared their performance. All the numerical results are presented in Table 3. We investigated the influence of the relaxation parameter v on TRE, comparing TRE for the full graph representation as v is gradually changed. The results presented in Figure 4 show that the best performances of TRE of 1.16mm were achieved for $v = 0.15$ and $v = 0.1$. Further relaxation of the graph deteriorates TRE for the Dir-Lab dataset. The visualisation of the registration results for our method, for the most challenging Dir-Lab case 8 are shown in Figure 5. Distribution of the landmark error for the same case after registration using our method is presented in Figure 6.

3.6 Sliding Motion

To explicitly show how the proposed method deals with sliding motion, we calculated the maximum shear stretch of the deformation field, which can be used as a measure of sliding motion. Following the method presented in Ref. 49, we calculated the gradients of the deformation field with respect to x , y and z , similarly to calculating the Jacobian of the deformations. Subsequently,

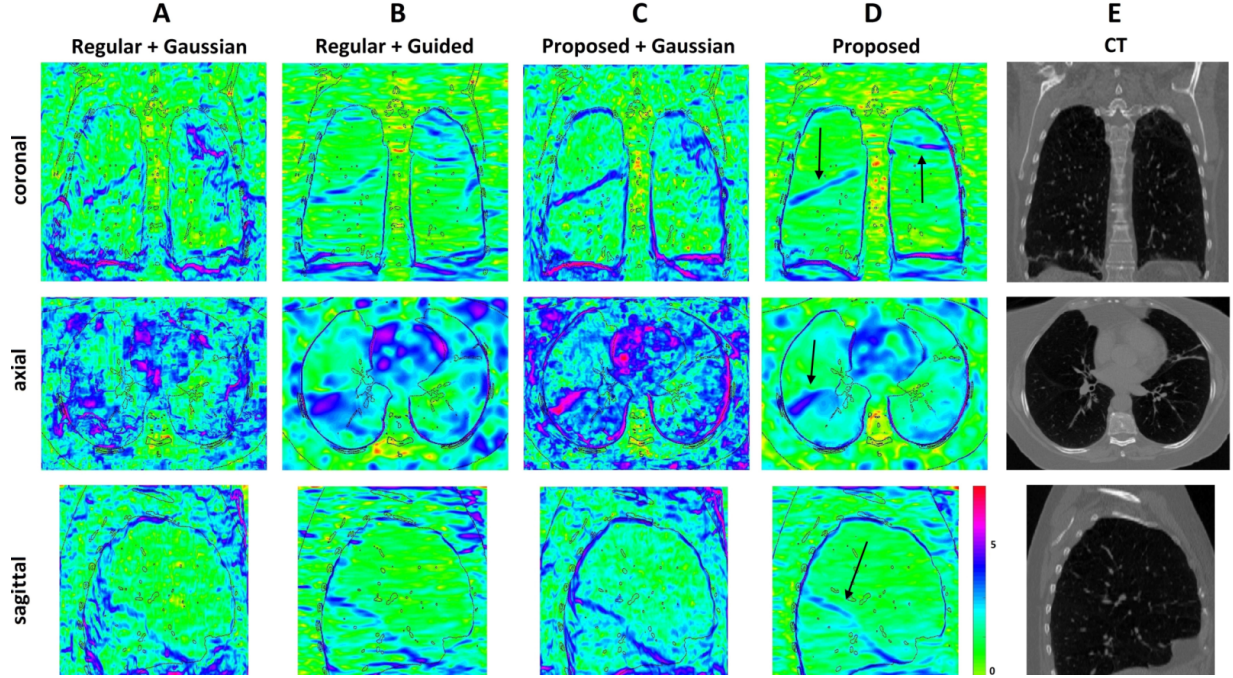


Fig 7 Example of maximum shear stretch calculated for the most challenging case 8 shown in coronal, axial and sagittal view. We compared the results for different variants of the proposed method, presenting the regular grid with Gaussian filtering on the column A, the regular grid representation with guided image filtering on the column B, supervoxels-based image representation with Gaussian filtering on the column C and on the column D the proposed method, where supervoxel image representation with guided image filtering was used. Fissures, pointed by black arrows between the lung lobes, indicate a fair amount of sliding motion taking place inside the lungs. On the last column E the corresponding anatomical CT images are shown, from which the main contours were propagated on the previous images.

we decomposed the deformation field gradients into eigenvalues and found their maximum and minimum principal stretch components, γ_1 , γ_3 , respectively. Then the maximum shear stretch is defined as: $\gamma_{max} = \frac{1}{2}(\gamma_1 - \gamma_3)$. The visualization of the maximum shear stretch distribution in logarithmic scale is shown in the example of case 8 in Figure 7. The highest value of shear stretch $\gamma_{max} > 5$ almost everywhere at lungs borders indicates a considerable sliding motion at borders between the lungs and the rib cage and also the diaphragm and the liver. It is also clearly noticeable (pointed by black arrows in Figure 7) that the value of shear stretch between the lung lobes is equal or slightly below 5, which implies interlobar sliding motion.

3.7 Comparison to Other Methods

We compared our method against other, most closely related methods in the field, based on the TRE reported for the Dir-Lab data set. These are: continuous optimization-based Demons method enhanced with bilateral-filtering (BLF¹⁶), dense displacement sampling method using belief propagation on a minimum spanning tree (deeds¹⁸) and supervoxel-based with belief propagation method applied over a minimum spanning tree (SBP²⁰), where the minimum spanning tree is extracted from a graph created by the adjacent supervoxels and the optimization performed over a number of layers of supervoxels. Our proposed method achieved a mean TRE of 1.16mm, compared to 1.96mm for BLF, 1.43mm for deeds and 1.23mm for SBP. For all of the cases the proposed method achieved the lowest TRE, except case 8, where the best result has been achieved with the SBP method. The numerical results are presented in Figure 8. Results achieved for our method show statistically significant improvement compared to those using BLF and deeds (p-value <0.05). However, no statistical significance was found between the results for SBP.

4 Discussion and Conclusions

We have proposed a novel deformable image registration method, which successfully combines compact image representation via supervoxel clustering with an efficient graph cuts-based optimization method and guided image filtering for providing locally continuous displacement field. We have shown the justification for the chosen approach by presenting at first the results for regular grid image representation variant, in which we extracted cubic supervoxels. Even though the compactness parameter m for SLIC was set to a relatively high value (4000), it is possible that some of the the supervoxels were imperfect cubes. Nevertheless, it only emphasises the fact that the perfectly regular cubic representation would further deteriorate the results. For the regular grid

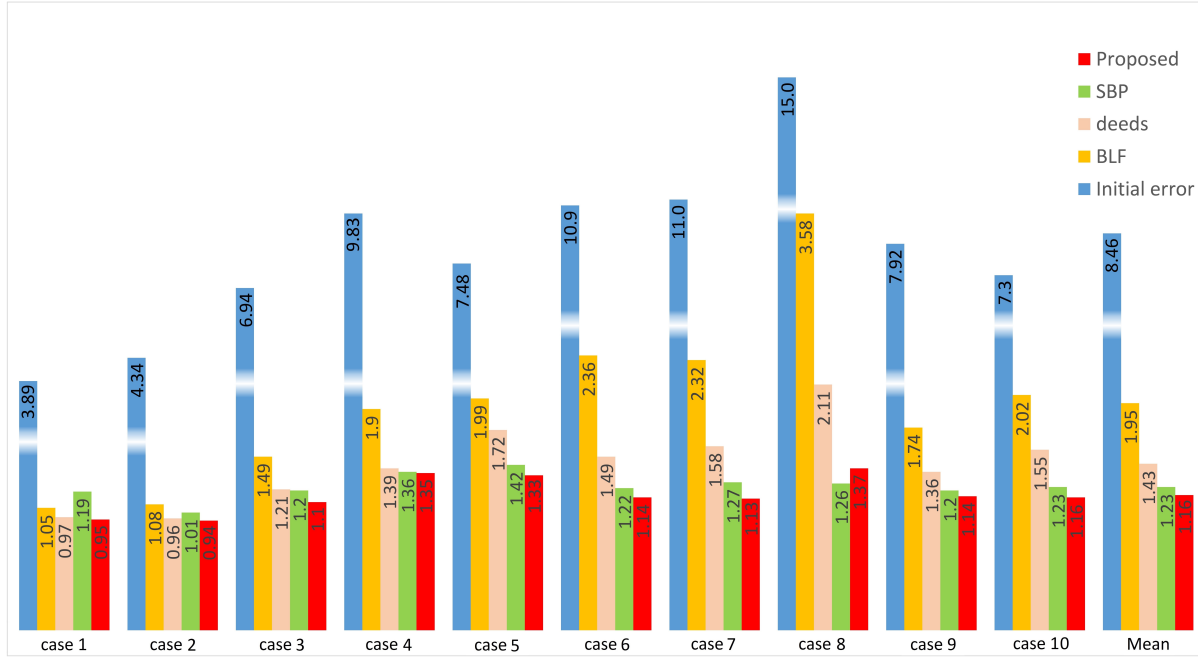


Fig 8 Comparison of the proposed method with the closest implementations based on the mean TRE for Dir-Lab data set. For each case, the TRE before registration (Initial) and after registration are shown: our proposed method, after bilateral-filtering combined with Demons (BLF¹⁶), after the dense displacement sampling method based on belief propagation on a minimum spanning tree (deeds¹⁸) and after supervoxel-based belief propagation (SBP²⁰). Mean results show that our proposed method outperforms all of the compared methods on average and in all individual cases apart from case 8.

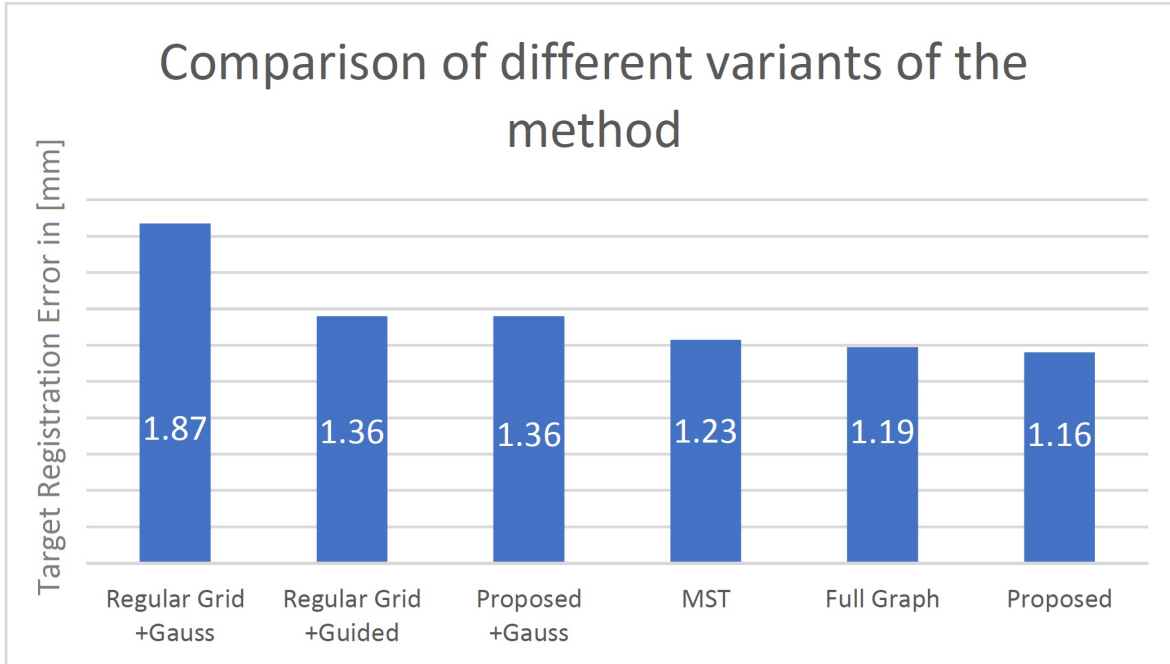


Fig 9 Comparison of different variants of the method based on TRE in mm.

image representation with Gaussian filtering we achieved TRE of 1.87mm, which performed worst among all of the compared methods and seems to be the least suitable configuration for the application. After changing the image representation to supervoxels and maintaining the Gaussian filtering, TRE is reduced to 1.36mm, which makes it the most noticeable change in TRE among all of variants of the method. It clearly shows the superiority of supervoxel-based image representation over a regular grid. The relative difference between them, compared to the initial TRE of 15mm, may not seem like large; however visual assessment of the magnitude of the displacement field from Figure 3 illustrates how much this is advantageous in practice. The displacement field for regular grid is heavily distorted at lung borders and the shape of the lungs is not recognisable, even though we would expect the lungs to be locally consistent in the movement. It clearly indicates that the supervoxel-based image representation has superior properties over its regular grid counterpart. Surprisingly, the supervoxel image representation with Gaussian filtering variant has achieved the same TRE (1.36mm), as for the regular grid method with guided image filtering. It suggests that the guided image filtering method is more suitable than Gaussian filtering in this application for image registration task and plays a role in additional structure oriented regularization, which can partially compensate for over-segmentation using supervoxels. These reasons strongly support the approach proposed in our method. Good results for the minimum spanning tree-based version were expected, based on the previously reported results in the literature.^{18,20} They position themselves well between the regular grid/Gaussian filtering versions and the proposed method. Coincidentally the same average TRE (1.23mm) is achieved in our test and the one reported in Ref. 20, however the specific cases of the dataset achieved different TREs. The second best results were achieved for the proposed version in a full graph representation (TRE of 1.19mm), which were further improved by its relaxed variant (TRE 1.16mm). The relaxed graph representation provided improved

or at least equally good results for all cases except for case 10. In our approach we decided to tune the parameters using just one of the cases of the dataset - case 8, which is the most challenging and has the largest initial misalignment. Such an approach is more robust to over-fitting the parameters. As a drawback it is possible that such an approach will result in suboptimal parameters settings for one (or more) of the cases. A larger dataset could potentially result in more accurate parameter tuning. We visualize the results for different variants of the proposed method in Figure 9, showing the gradual improvement of the results. As an additional experiment we investigated the influence of the intensity term in the piecewise smoothness term of the energy formulation in Equation 8. The substitution of the proposed linear intensity term with a quadratic one resulted in TRE of 1.22mm, compared to 1.16mm for the linear term. This experiment supports our proposed formulation of the energy in Equation 8. In terms of the time, the performance of the proposed method in its current suboptimal and non-parallelized Matlab implementation varied from 63 min to 121 min per pairwise registration, depending on the Dir-Lab case.

In our comparison to other methods we used the closest methods in approach, at the same time considered as the-state-of-the-art in medical image registration. The authors of the work from Ref. 15 did not use the Dir-Lab dataset for their validation. We compared our results with Ref. 16, which shares with Ref. 15 the log-Demons optimization strategy. In referring to the work from Ref. 28, its memory requirements were far beyond our available resources. The authors of Ref. 29, which can be seen as an extension of work from Ref. 28, reported the computation time for a single 3D registration to be up to 24h. In our case it was between 63 and 121 min, depending on the volume size. It was just slightly over 1h for images having similar volume with the ones from Ref. 29. The methods proposed in Refs. 31 and 32 have been presented for 2D cases only. They could potentially be extended to 3D cases, however this would have required significant code

development, which was beyond the scope of our work. The best performing method on the Dir-Lab dataset at this moment is Ref. 50 with TRE of 0.94mm on average. This method and other top performing methods on the Dir-Lab dataset use segmentations and perform registration only inside the masks, as all the landmarks are placed inside the lungs. Such an approach makes the methods unsuited for direct comparison.

The proposed method follows the trend noticeable in image registration field where discrete optimization-based methods perform very well and in some cases favorably compared to their continuous optimization-based counterparts. Our proposed framework contributes to the field addressing the aims summarized in Section 1.3. The proposed method creates a promising framework for more advanced applications, such as multi-modal image fusion, where images originating from different modalities could be combined together. In future we intend to apply the method for hyperpolarized gas MRI to CT image registration using proton MRI as an intermediate registration step. Another interesting application of the method might be a joint lung registration and lobe segmentation. The borders between the lobes look like if they could be well distinguishable based on the deformation field, as show in Figure 7. In our opinion, the method may be applicable to other medical image registration where sliding motion appears, such as multi-modal lung registration, liver or cardiac motion correction. Supervoxel-based image representation appears to be still a very promising method, which however, has not yet been fully explored in image registration.

Disclosures

The authors have no relevant financial interests in this article and no potential conflicts of interest to disclose.

Acknowledgments

AS, BP and VG gratefully acknowledge funding from the CRUK & EPSRC Cancer Imaging Centre in Oxford. AS would like to acknowledge Prof. M. Pawan Kumar and Prof. Olga Veksler for their advice and suggestions. AH acknowledges the support of the Research Council UK Digital Economy Programme grant number EP/G036861/1 (Oxford Centre for Doctoral Training in Healthcare Innovation) and the CAPES Foundation, process No BEX 0725/12-9.

References

- 1 B. Fulkerson, A. Vedaldi, and S. Soatto, “Class segmentation and object localization with superpixel neighborhoods,” *IEEE Int. Conf. Comput. Vis.* , 670–677 (2009).
- 2 H. Kim, S. Lee, D. Lee, *et al.*, “Real-time human pose estimation and gesture recognition from depth images using superpixels and svm classifier,” *Sensors* **15**(6), 12410–12427 (2015).
- 3 C. L. Zitnick and S. B. Kang, “Stereo for image-based rendering using image over-segmentation,” *Int. J. Comput. Vision* **75**, 49–65 (2007).
- 4 B. Liu, S. Gould, and D. Koller, “Single image depth estimation from predicted semantic labels,” *Comput. Vis. Pattern Recognit.* , 1253–1260 (2010).
- 5 J. A. Schnabel, M. P. Heinrich, B. W. Papież, *et al.*, “Advances and challenges in deformable image registration: From image fusion to complex motion modelling.,” *Med. Image Anal.* **33**, 145–148 (2016).
- 6 S. Shaker, A. Dirksen, L. Laursen, *et al.*, “Volume adjustment of lung density by computed tomography scans in patients with emphysema.,” *Acta Radiologica* **4** (2004).

- 7 D. F. Pace, S. R. Aylward, and M. Niethammer, “A locally adaptive regularization based on anisotropic diffusion for deformable image registration of sliding organs,” *IEEE Trans. Med. Imag* **32**(11), 2114–2126 (2013).
- 8 R. M. Kaus, T. Netsch, S. Kabus, *et al.*, “Estimation of Organ Motion from 4D CT for 4D Radiation Therapy Planning of Lung Cancer,” *Med. Image Comput. Comput.-Assist. Intervent. (MICCAI 2004)* **3217**, 1017–1024 (2004).
- 9 E. Weiss, K. Wijesooriya, S. V. Dill, *et al.*, “Tumor and normal tissue motion in the thorax during respiration: Analysis of volumetric and positional variations using 4D CT,” *Med. Image Comput. Comput.-Assist. Intervent. (MICCAI 2004)* **67**, 296307 (2007).
- 10 R. Castillo, E. Castillo, and T. Guerrero, “Ventilation from four- dimensional computed tomography: density versus Jacobian methods. ,” *Phys. Med. Biol.* **55**, 4661–4685 (2010).
- 11 T. Rohlfing, “Image similarity and tissue overlaps as surrogates for image registration accuracy: Widely used but unreliable.,” *IEEE Trans. Med. Imag* **31**(2), 153–163 (2012).
- 12 B. Horn and B. Schunck, “Determining optical flow.,” *Artificial Intelligence* **17**, 185–203 (1981).
- 13 A. Schmidt-Richberg, R. Werner, H. Handels, *et al.*, “Estimation of slipping organ motion by registration with direction-dependent regularization,” *Med. Image Anal.* **16**, 150–159 (2012).
- 14 H. Y. Baluwala, L. Risser, J. A. Schnabel, *et al.*, “Toward physiologically motivated registration of diagnostic CT and PET/CT of lung volumes.,” *Med. Phys.* **40**(2), 021903 (2013).
- 15 L. Risser, F. X. Vialard, H. Y. Baluwala, *et al.*, “Piecewise-diffeomorphic image registration: application to the motion estimation between 3D CT lung images with sliding conditions,” *Med. Image Anal.* **17**(2), 182–193 (2013).

- 16 B. W. Papież, M. P. Heinrich, J. Fehrenbach, *et al.*, “An implicit sliding-motion preserving regularisation via bilateral filtering for deformable image registration.,” *Med. Image Anal.* **18**, 1299–311 (2014).
- 17 J. Thirion, “Image matching as a diffusion process: an analogy with Maxwell’s demons,” *Med. Image Anal.* **2**(3), 243–260 (1998).
- 18 M. P. Heinrich, M. Jenkinson, M. Brady, *et al.*, “MRF-based deformable registration and ventilation estimation of lung CT.,” *IEEE Trans. Med. Imag.* **32**, 1239–1248 (2013).
- 19 M. P. Heinrich, H. Handels, and I. J. A. Simpson, “Estimating large lung motion in COPD patients by symmetric regularised correspondence fields,” *Med. Image Comput. Comput.-Assist. Intervent. (MICCAI 2015)* **9350**, 338–345 (2015).
- 20 M. P. Heinrich, I. J. Simpson, B. W. Papież, *et al.*, “Deformable image registration by combining uncertainty estimates from supervoxel belief propagation,” *Med. Image Anal.* **27**, 57–71 (2016).
- 21 Y. Boykov and G. Funka-Lea, “Graph cuts and efficient N-D image segmentation,” *Int. J. Comput. Vis.* **70**(2), 109–131 (2006).
- 22 P. F. Felzenszwalb and D. P. Huttenlocher, “Efficient graph-based image segmentation.,” *Int. J. Comput. Vis.* **59**(2), 167–181 (2004).
- 23 B. Glocker, N. Komodakis, G. Tziritas, *et al.*, “Dense image registration through MRFs and efficient linear programming,” *Med. Image Anal.* **12**(6), 731–741 (2008).
- 24 D. Cobzas and A. Sen, “Random walks for deformable image registration,” *Med. Image Comput. Comput.-Assist. Intervent. (MICCAI 2011)* , 557–565 (2011).

- 25 A. Raj and R. Zabih, “A graph cut algorithm for generalized image deconvolution,” *IEEE International Conference on Computer Vision (ICCV’05)* **2**, 1048–1054 (2005).
- 26 Y. Boykov and V. Kolmogorov, “An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision,” *IEEE Trans. Pattern Anal. Mach. Intell.* **26**, 1124 – 1137 (2004).
- 27 J. Kim, V. Kolmogorov, and R. Zabih, “Visual correspondence using energy minimization and mutual information,” *IEEE Int. Conf. Comput. Vis.* , 1033–1040 vol.2 (2003).
- 28 T. W. H. Tang and A. C. S. Chung, “Non-rigid image registration using graph-cuts,” *Med. Image Comput. Comput.-Assist. Intervent. (MICCAI 2007)* **10**(Pt 1), 916–924 (2007).
- 29 R. W. K. So, T. W. H. Tang, and A. C. S. Chung, “Non-rigid image registration of brain magnetic resonance images using graph-cuts,” *Pattern Recognit.* **44**(10-11), 2450–2467 (2011).
- 30 H. Lombaert, Y. Sun, and F. Cheriet, “Landmark-based Non-rigid Registration via Graph Cuts,” *Lec. Notes Comput. Sc.* (2007).
- 31 R. W. K. So and A. C. S. Chung, “Non-rigid image registration by using graph-cuts with mutual information,” *IEEE Image Proc.* **Sept**(1522-4880), 4429–4432 (2010).
- 32 R. W. K. So and A. C. S. Chung, “Learning-based non-rigid image registration using prior joint intensity distributions with graph-cuts,” *IEEE Image Proc.* **2**, 717–720 (2011).
- 33 R. Achanta, A. Shaji, K. Smith, *et al.*, “SLIC superpixels compared to state-of-the-art superpixel methods,” *IEEE Trans. Pattern Anal. Mach. Intell.* (2012).
- 34 R. Sedgewick, “Algorithms in c, part 5: Graph algorithms,” *Addison-Wesley, Reading, Massachusetts* (2001).

- 35 K. He, J. Sun, and X. Tang, “Guided Image Filtering,” *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(6), 1397–1409 (2013).
- 36 A. Szmul, B. W. Papież, R. Bates, *et al.*, “Graph Cuts-Based Registration Revisited: A Novel Approach for Lung Image Registration Using Supervoxels and Image-Guided Filtering,” *Comput. Vis. Pattern Recognit. Workshops (CVPRW)* , 592–599 (2016).
- 37 A. Vedaldi and S. Soatto, “Quick shift and kernel methods for mode seeking,” *ECCV Lec. Notes Comput. Sc.* **5305**(4), 705–718 (2008).
- 38 V. Machairas, M. Faessel, D. Cárdenas-Peña, *et al.*, “Waterpixels,” *IEEE Trans. Image Proc.* **24**(11), 3707–3716 (2015).
- 39 M. P. Heinrich, M. Jenkinson, M. Bhushan, *et al.*, “MIND: modality independent neighbourhood descriptor for multi-modal deformable registration.,” *Med. Image Anal.* **16**, 1423–1435 (2012).
- 40 J. Ruhaak, S. Heldmann, T. Kipshagen, *et al.*, “Highly accurate fast lung CT registration,” *SPIE Med. Imag.* **86690Y** (2013).
- 41 J. Vandemeulebroucke, O. Bernard, S. Rit, *et al.*, “Automated segmentation of a motion mask to preserve sliding motion in deformable registration of thoracic CT,” *Med. Phys.* **39** (2012).
- 42 Y. Boykov, O. Veksler, and R. Zabih, “Fast approximate energy minimization via graph cuts,” *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(11), 1222–1239 (2001).
- 43 D. Rueckert, L. I. Sonoda, C. Hayes, *et al.*, “Nonrigid registration using free-form deformations: application to breast MR images.,” *IEEE Trans. Med. Imag* **18**, 712–21 (1999).
- 44 C. Tomasi and R. Manduchi, “Bilateral Filtering for Gray and Color Images,” *Int. J. Comput. Vis.* , 839–846 (1998).

- 651 45 B. W. Papież, J. Franklin, M. P. Heinrich, *et al.*, “Liver Motion Estimation via Locally
652 Adaptive Over-Segmentation Regularization,” *Med. Image Comput. Comput.-Assist. Inter-*
653 *vent. (MICCAI 2015)* **9351**, 427–434 (2015).
- 654 46 M. Chen, W. Lu, Q. Chen, *et al.*, “A simple fixed-point approach to invert a deformation
655 field,” *Med. Phys.* **35**(1), 81–88 (2008).
- 656 47 R. Castillo, E. Castillo, R. Guerra, *et al.*, “A framework for evaluation of deformable image
657 registration spatial accuracy using large landmark point sets,” *Phys. Med. Biol.* **54**(7), 1849–
658 1870 (2009).
- 659 48 J. Ashburner, “A fast diffeomorphic image registration algorithm,” *NeuroImage* **38**(1), 95–
660 113 (2007).
- 661 49 R. E. Amelon, K. Cao, J. M. Reinhardt, *et al.*, “A measure for characterizing sliding on lung
662 boundaries,” *Annals of Biomedical Engineering* **42**(3), 642–650 (2014).
- 663 50 L. Knig and J. Ruhaak, “A fast and accurate parallel algorithm for non-linear image reg-
664 istration using normalized gradient fields,” *IEEE International Symposium on Biomedical*
665 *Imaging* (2014).

List of Figures

- 1 Diagram presenting the workflow of the proposed method. We start from extracting supervoxels and creating a graph based on them. Then we estimate a displacement for each supervoxel using the graph cuts optimization method. The guided image filtering is first applied to the estimated update of the displacement field, similar to fluid-like regularization in Demons, and then to the update combined with the previously estimated displacement field, similar to diffusion-like regularization in Demons. We then estimate inverse transformations and combine them with forward ones. Double rectangles emphasize the symmetric nature of the method, where forward and backward transformation from the target to the moving image are estimated at the same time.
- 2 Example of the SLIC algorithm clustering of the peak inhale 2D lung CT image. In (a) the original CT image is shown, (b) shows the corresponding supervoxel representation, with the borders of the supervoxels drawn in blue. The graph extracted based on the clustering is shown in figure (c), with the edges connecting adjacent supervoxels in red. The relaxed graph representation is shown in figure (d), whereas the minimum spanning tree-based simplification of the graph is shown in figure (e). In all of the images, apart from the first one (a), the contrast has been adjusted for better visualisation of supervoxels and graphs. For illustrative purposes, we use 2D superpixels and present them on a 2D slice. However, in the framework we extract supervoxels from 3D volumes.

- 3 Magnitude of the displacement field in the coronal view for different image representations and filtering methods. On the upper row we show the results for regular grid image representation, while on the lower row for supervoxel-based image representation. The left column presents the results for Gaussian filtering and the right for guided image filtering. The best results were achieved with the proposed configuration shown in the lower right corner.
- 4 Influence of the relaxation parameter ν on TRE in [mm] and standard deviation for the DIR-Lab dataset. The scale in x axis is non-linear
- 5 Registration results for the proposed method for the most challenging Dir-Lab case 8 in axial, coronal and sagittal view. In the left column an overlay of full inhale (green) and full exhale (magenta) before applying the registration is presented. In the middle column the same slices after performing the proposed registration are shown. The initial misalignment, visible in the left column, is almost completely eliminated by the proposed registration method. In the right column the magnitude of the displacement field in mm is presented.
- 6 Visualization of the landmark error distribution for the most challenging case 8 after registration using our method in all three views with the lung mask in the background. Only landmarks with an error higher than 1.5mm are displayed. The size of the points corresponds to the landmark error.

7 Example of maximum shear stretch calculated for the most challenging case 8
shown in coronal, axial and sagittal view. We compared the results for different
variants of the proposed method, presenting the regular grid with Gaussian filtering
on the column A, the regular grid representation with guided image filtering on the
column B, supervoxels-based image representation with Gaussian filtering on the
column C and on the column D the proposed method, where supervoxel image
representation with guided image filtering was used. Fissures, pointed by black
arrows between the lung lobes, indicate a fair amount of sliding motion taking
place inside the lungs. On the last column E the corresponding anatomical CT
images are shown, from which the main contours were propagated on the previous
images.

8 Comparison of the proposed method with the closest implementations based on the
mean TRE for Dir-Lab data set. For each case, the TRE before registration (Initial)
and after registration are shown: our proposed method, after bilateral-filtering com-
bined with Demons (BLF¹⁶), after the dense displacement sampling method based
on belief propagation on a minimum spanning tree (deeds¹⁸) and after supervoxel-
based belief propagation (SBP²⁰). Mean results show that our proposed method
outperforms all of the compared methods on average and in all individual cases
apart from case 8.

9 Comparison of different variants of the method based on TRE in mm.

List of Tables

- 1 Comparison of the proposed method against regular grid image representation, based on TRE in [mm] and standard deviation for the Dir-Lab dataset. The best results are shown in bold that are consistently best for the proposed method using supervoxels and guided image filtering
- 2 Comparison of the proposed method against its variants based on fraction of negative determinant of the Jacobian of deformations, which measures the fraction of voxels for which a one-to-one mapping is violated, mean of the determinant of the Jacobian of deformation, showing the change in volume and standard deviation of the determinant of the Jacobian of deformations, which can be seen as a measure of complexity of the deformations, for the Dir-Lab dataset.
- 3 Comparison of the proposed method applied over the full graph without the relaxation criterion, after applying the proposed graph relaxation and after relaxing the graph to a minimum spanning tree (MST), based on TRE and standard deviation for Dir-Lab dataset. Best results are shown in bold.