

Supervoxels for graph cuts-based deformable image registration using guided image filtering

Adam Szmul
Bartłomiej W. Papież
Andre Hallack
Vicente Grau
Julia A. Schnabel

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 and Julia A. Schnabel^{a,b}

^aUniversity of Oxford, Institute of Biomedical Engineering, Department of Engineering Science, Oxford, United Kingdom

^bKing's College London, Department of Biomedical Engineering, School of Biomedical Engineering and Imaging Sciences, London, United Kingdom

Abstract. We propose combining a supervoxel-based image representation with the concept of graph cuts as an efficient optimization technique for three-dimensional (3-D) deformable image registration. Due to the pixels/voxels-wise graph construction, the use of graph cuts in this context has been mainly limited to two-dimensional (2-D) 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 3-D 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 leads to the observation that our approach compares very favorably with state of the art methods in continuous and discrete image registration, achieving target registration error of 1.16 mm on average per landmark. © 2017 SPIE and IS&T [DOI: [10.1117/1.JEI.26.6.061607](https://doi.org/10.1117/1.JEI.26.6.061607)]

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

Paper 170333SS received Apr. 25, 2017; accepted for publication Sep. 7, 2017; published online Oct. 4, 2017.

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 estimation,² and 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 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 three-dimensional (3-D) 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 for an assessment of the same structures in 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 breathing,⁶ 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 in the literature to as “sliding motion”. It is characterized by discontinuities at the surface of the lungs in tangential components of the displacement field along the direction normal to the surface, whereas components in the direction tangential to the surface inside and outside of the lungs, as well as the components normal to them across the surface, 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 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. To accurately assess the same structures, a very careful image registration needs to be performed to allow for direct side-by-side comparison. Practical application 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

*Address all correspondence to: Adam Szmul, E-mail: adam.szmul@eng.ox.ac.uk

painting), taking the different breathing states into account. Another application is the 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 subvoxel 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 toward the fixed one. This process can be shown as an energy minimization procedure of a similarity measure of the images, $\text{Sim}\{I_{\text{fix}}, I_{\text{mov}}[\mathbf{u}(\mathbf{x})]\}$, with $I_{\text{mov}}[\mathbf{u}(\mathbf{x})]$ 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 purely on similarity measure, which matches spatial locations based on the chosen measure, might result in 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 $\text{Reg}[\mathbf{u}(\mathbf{x})]$, which penalizes implausible deformations, is introduced to further constrain the transformation model. The exact definition of the $\text{Reg}[\mathbf{u}(\mathbf{x})]$ 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

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

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 this method is that it requires accurate lung segmentation for calculating normal and tangential direction vectors. Another limitation 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 to CT image registration.¹⁴ A piecewise diffeomorphism, as an extension to large deformation diffeomorphic metric mapping, was proposed in Ref. 15, whereas bilateral filtering, as the 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 was presented in Ref. 18 in which the optimization problem was posed on a minimum spanning tree (MST) 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 field method was introduced in Ref. 19 for images from patients with chronic obstructive pulmonary disease and a supervoxel-based belief propagation algorithm in Ref. 20. For the latter, the optimization has also been posed on a MST, 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 the lungs. This will break up the regularization components that we would expect between those regions. In contrast, a graph such as the one proposed here preserves all relevant connections.

1.2 Markov Random Fields and Graph Cuts

To use a discrete solver for deformable image registration, the problem is usually formulated in terms of MRFs, which 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 neighboring nodes. The optimization is posed on an undirected graph that is built by a set of nodes P . In such a form, the optimization is the process of finding a labeling f for the set P that minimizes the energy $E(f)$, consisting of the sum of two terms

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

where $E_{\text{data}}(f)$ represents the disagreement between the labeling f and the observed data, whereas an α -weighted piecewise smoothness term $E_{\text{smooth}}(f)$ forces nodes connected by an edge to have similar labels. Depending on the application, the labels can represent different properties, for instance objects or background in segmentation²¹ or a depth map in stereovision.²² Such a formulation can also be applied to deformable image registration, where the labels to be assigned represent predefined displacement vectors $[u_x, u_y, u_z]$. For this application, only a limited number of displacements are considered; thus this formulation requires a transformation model that transfers the displacement field from a discrete domain into a continuous one. However, for a 3-D application, the number of labels consists of thousands of allowed displacements.

Methods capable of solving such a formulation have attracted researchers since its introduction. However, the lack of efficient optimization algorithms limited their

application for deformable image registration. An approach using dynamic programming as an optimization method, drop,²³ has been successfully applied for brain image registration. The main limitation of this method is that it estimates the best displacements in the x , y , and z directions independently. Moreover, it does it in an iterative way, which renders the method similar to the continuous-optimization-based approaches and therefore potentially vulnerable to be trapped in a local minimum. These restrictions limit the advantage of the discrete optimization-based approach, which allows for a dense displacement sampling and for some methods to find at least a strong local minimum. In Ref. 24, the authors proposed using a random walks algorithm for deformable image registration. Both methods from Refs. 18 and 20, presented before, use the belief propagation algorithm as the optimization method to estimate displacement field.

Another example of the optimization method is graph cuts. Although the method has been widely used in many fields of computer vision, such as image segmentation,²¹ denoising,²⁵ stereovision,²⁶ and motion detection,²⁷ its application to image registration has been limited. A voxel-wise deformable registration method applying graph cuts was proposed in Ref. 28 for both 2-D and 3-D brain images. An obvious limitation of the method is the size of the graph on which the problem is posed. While for 2-D images the problem is still tractable, in 3-D cases the size of the graph with every voxel is represented by a node might be much larger. This, along with the high number of possible displacements, makes 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 3-D cases of up to 24 h. A landmark-based graph cuts algorithm was proposed for medical image registration in Ref. 30. In that work, a penalty cost was added to the data term for a node on which a landmark was 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 the 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 supervoxels. When each of the supervoxels consists of n_s voxels, then the number of supervoxels $N = \frac{n}{n_s}$, and the worst-case computational complexity of our formulation is expressed as $O(N^2)$. Subsequently, the number of displacements to be estimated is reduced from the number of voxels in the image to the number of extracted supervoxels, which is significantly lower. The use of supervoxels has certain limitations. Small supervoxels diminish the advantage of the reduction of the number of nodes in the graph (in the limit, the number of supervoxels can become the number of voxels). On the other hand, when the

supervoxels are large, they might not be able to accurately model the variations in displacement. In the process of finding the optimum displacements, we apply the graph cuts method and later use an edge-preserving filtering method, namely guided image filtering,³⁵ to convert the displacement field derived from the sparse image representation into a piecewise continuous displacement field. The proposed method has been initially presented in Ref. 36. The paper has been deeply extended and rewritten. In particular, here we present the method in full detail. Second, we provide a comprehensive evaluation of different variants of the method, justifying the chosen approach. Third, we compare the method against state of the art methods including a regular grid (RG) image representation and investigate the influence of filtering methods. Finally, we also show how the method and its variants handle the sliding motion modeling. The closest method in the field to the proposed is the one presented in Ref. 20. Even though both methods share some common features, such as the supervoxel approach to the image parametrization or discrete optimization, closer analysis reveals that the methods present completely different approaches to image registration. In our method, we use a multiresolution approach to extract supervoxels, whereas in Ref. 20 the supervoxels are extracted from the original, unprocessed images. In our method, at every resolution level we extract supervoxels with different relative sizes, starting 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 Sec. 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 in the size of the 3-D deformable image registration problem, making it efficiently solvable via the graph cuts method using supervoxels.
2. To introduce a graph relaxation criterion based on a relation between intensities of adjacent supervoxels, which improves the 3-D 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 to those that do not use supervoxel image representation, when compared on a publicly available dataset.

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

2 Methods

In this section, we present the proposed method for deformable image registration, using supervoxel-based image representation and graph cuts as an optimization. An overview of the proposed method is shown in Fig. 1, with every component addressed individually in the following sections: in Sec. 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 Sec. 2.2. A detailed explanation of the graph construction and energy function formulation used for the initial displacement field estimation is presented in Sec. 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 Sec. 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 RG 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, which group pixels with similar appearance within a local neighborhood, were proposed. In the literature, many approaches have been proposed to perform image clustering, including mean-shift methods,²² gradient-based methods,³⁷ and the watersheds method.³⁸ However, probably the most popular image clustering method is the 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 Υ 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 $d_I = \|I(\mathbf{x}) - I(\mathbf{c})\|$. The distance Υ is defined as $\Upsilon = [(d_e)^2 + (\frac{d_I}{S})^2 m^2]^{1/2}$, where m is a parameter that controls the compactness of supervoxels.

2.2 Similarity Measure

The simplest similarity measures relying purely on intensities are not well suited for CT lung registration, mainly due to changes in lung density during breathing, which result 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 using the Modality Independent Neighbourhood Descriptor (MIND),³⁹ which was 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

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

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 $\text{Var}(I, \mathbf{x})$ in the denominator in Eq. 3 stands for the estimation of intensity

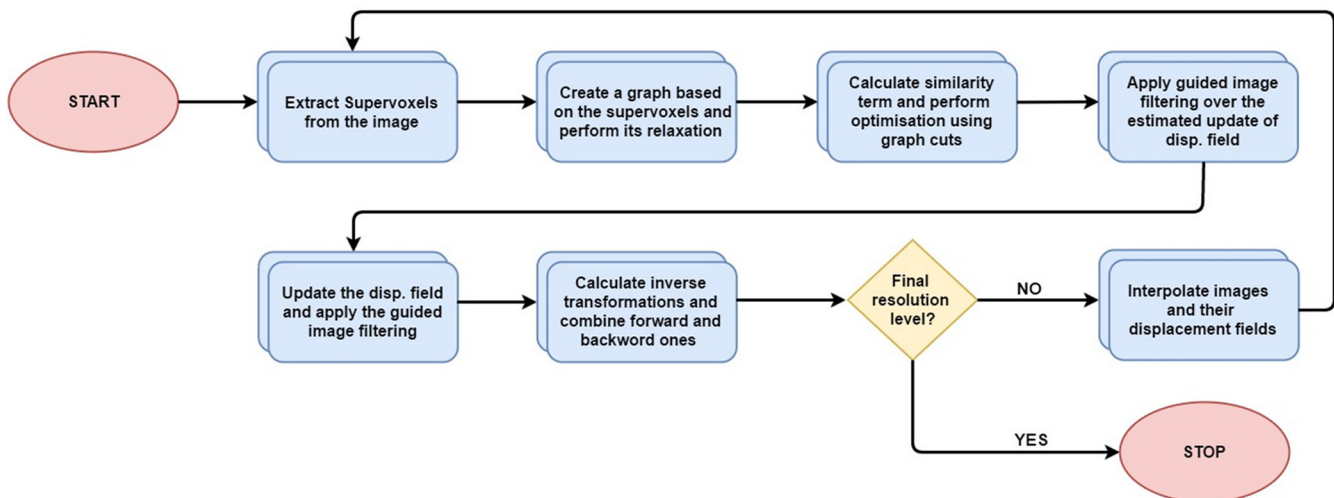


Fig. 1 Diagram presenting the workflow of the proposed method. We start from extracting supervoxels and creating a graph based on them. Then we estimate a displacement for each supervoxel using the graph cuts optimization method. The guided image filtering is first applied to the estimated update of 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 transformations from the target to the moving image are estimated at the same time.

distribution in the neighborhood of a voxel \mathbf{x} and is represented as the variance in intensities within a neighborhood N . For more details of the method, we refer the interested reader to Ref. 39. The main advantage of the MIND over single value intensity-based similarity measures is that it provides a descriptor that 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.

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

where $\overline{I_{\text{mov}}(\mathbf{x}_p)}$ and $\overline{I_{\text{mov}}(\mathbf{x}_q)}$ stand for mean intensity values of all voxels \mathbf{x}_p and \mathbf{x}_q contained within the supervoxels represented by nodes p and q , respectively, calculated for the moving image. The image intensities are initially normalized 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 nonzero edges remaining in the graph, the overall complexity of the graph layout is reduced, the optimization process is

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 at the same time preserving the smooth deformation field inside them, we propose using 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

performed with a better rate of the convergence. An example of such clustering and a graph created over the image, as well as its relaxed form, is shown in Fig. 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 presegmentation of the image, akin to lung masking, which is a common practice in the field of lung registration.^{15,40,41}

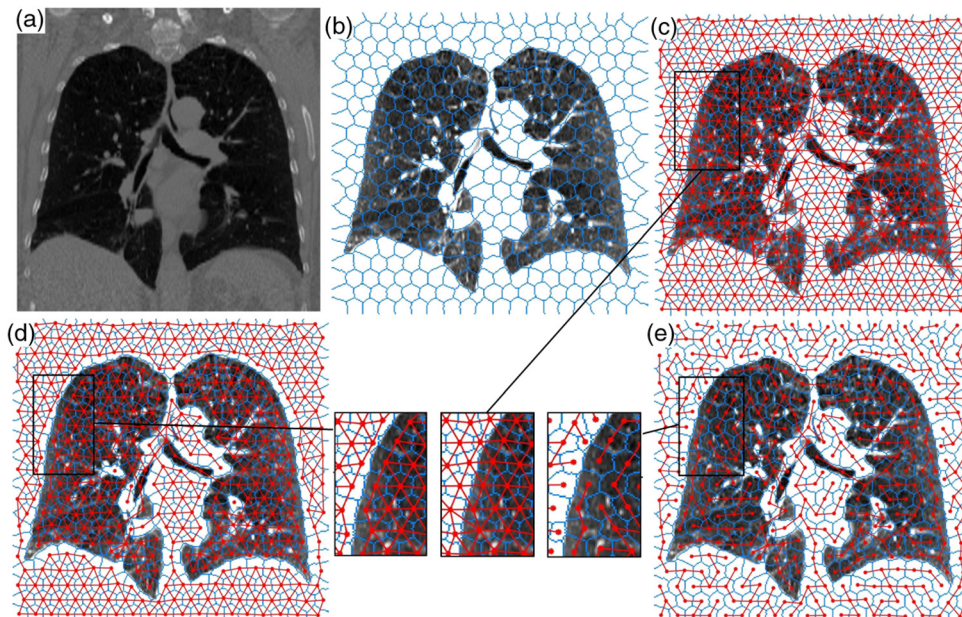


Fig. 2 Example of the SLIC algorithm clustering of the peak inhale 2-D lung CT image. (a) shows the original CT image, and (b) shows the corresponding supervoxel representation, with the borders of the supervoxels drawn in blue. The graph extracted based on the clustering is shown in figure (c), with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

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

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

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 (3-D image pixels) for each of the label, representing a displacement. Such a 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 neighboring nodes p and q , defined as

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

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

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

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

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

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

2.4 Guided Image Filtering

The labeling procedure, described in the previous section, results in all supervoxels being 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 the 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 with B-splines⁴³ is used as an interpolation method. The properties of B-splines, however, do not allow us to model directly 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. We use guided image filtering,³⁵ where an output image I_{out} is a linear combination of an input image I_{in} and an image I_g is used for guidance

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

where $\mathbf{W}_{\mathbf{y}}$ is a filter kernel calculated for the guide image I_g in a neighborhood $\mathcal{N}(\mathbf{x})$ 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} and cov_{I_g} are the mean and covariance of the guidance image I_g calculated in a neighborhood $\mathcal{N}(\mathbf{x})$, \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

To enhance the registration process as well as more accurately model the complex nature of the lungs during the breathing process, we propose applying 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 the

image toward the target and vice versa. Such a formulation not only compensates for some misregistrations 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

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

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

where $\mathbf{u}_{\text{mov} \rightarrow \text{fix}}$ and $\mathbf{u}_{\text{fix} \rightarrow \text{mov}}$ stand for the estimated deformation fields transforming I_{mov} to I_{fix} and 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 Sec. 3.1, we describe the dataset that was used for the evaluation purpose. The parameter setting is described in Sec. 3.2, and in Sec. 3.3, we explore the influence of the image representation by performing experiments on RG image parameterization. This is continued in Sec. 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 Sec. 3.5, we investigate the influence of the relaxation parameter, and in Sec. 3.6 we show how the method deals with sliding motions. Finally, in Sec. 3.7 we compare our method against others that are 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,^{47,48} which consists of 10 cases of dynamic 3-D CT volumes covering the full breathing cycle (also called four-dimensional CT) of patients suffering from esophageal or lung cancer. Patient identifiers were removed in accordance with an institutional review board 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 0.75 to 1.13 mm.⁴⁷ Such an approach allows us to assess how the methods deal 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 five resolution levels, at each of them extracting supervoxels of approximate sizes of [50, 150, 400, 800, 500] voxels with a maximum allowed displacement of [3, 4.5, 4.5, 6, 4.5] voxels and a 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 neighboring voxels, with a range of $r = 1$. The parameters have been chosen heuristically, based on the 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 RG representation, setting SLIC's compactness parameter m to 4000 (in all the other experiments we used $m = 40$), achieving TRE of 1.36 mm, compared with 1.16 mm 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 RG

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

TRE in (mm) and standard deviation				
Case	RG + Gauss	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

variant results in 2.42 mm of TRE, whereas for the proposed method TRE was only 1.37 mm. In Fig. 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

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

where $u_x(x, y, z)$, $u_y(x, y, z)$, and $u_z(x, y, z)$ are the displacement fields estimated in the x , y , and z directions, respectively. In Table 2, we present the fraction of the negative determinant of the Jacobian of deformations, which shows the fraction of voxels for which a one-to-one mapping is violated,⁴⁹ the mean of the determinant of the Jacobian of deformations, showing the change in volume, and the standard deviation of the determinant of the Jacobian of deformations, all calculated inside the lungs for the Dir-Lab dataset. We can observe a trend that the variants using supervoxel image representation result in a lower fraction of negative determinant of the Jacobian of deformations, as well as lower standard deviation of the determinant of Jacobian of deformations.

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

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	RG	Proposed + Gauss	Proposed
Fraction of neg J.	8.5×10^{-3}	1.0×10^{-3}	1.8×10^{-3}	7.2×10^{-3}
Mean Jacobian	0.910	0.918	0.915	0.918
Std of Jacobian	0.329	0.191	0.215	0.162

the guided image filtering. When supervoxel-based image representation was used, the Gaussian filtering achieved TRE of 1.36 mm, whereas for the proposed, it was 1.16 mm. A similar trend was observed when we applied both filtering methods for the RG image representation, resulting in TRE of 1.87 and 1.36 mm 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 Fig. 3, where we present the proposed method with guided image filtering, the proposed method with Gaussian filtering, and RG with both filtering methods. The results presented in Table 2 show that guided image filtering provides a lower fraction of negative determinant of the Jacobian of deformations as well as a lower standard deviation of the determinant of the Jacobian of deformations,

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 MST, based on TRE and standard deviation for Dir-Lab dataset. Best results are shown in bold.

TRE 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

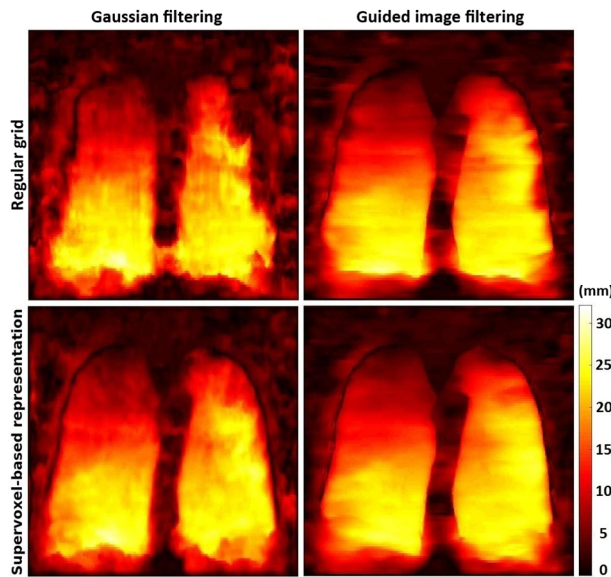


Fig. 3 Magnitude of the displacement field in the coronal view for different image representations and filtering methods. On the upper row we show the results for regular grid image representation, while on the lower row for supervoxel-based image representation. The left column presents the results 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.

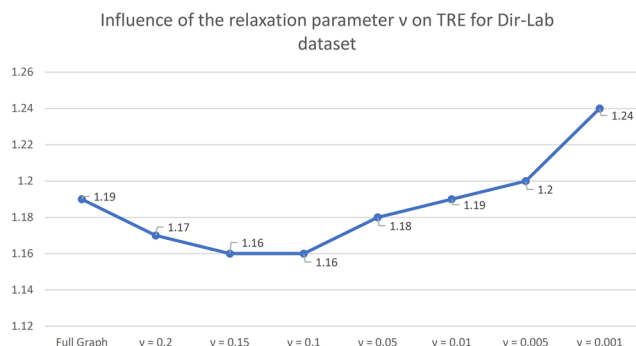


Fig. 4 Influence of the relaxation parameter v on TRE in (mm) and standard deviation for the DIR-Lab dataset. The scale in the x axis is nonlinear.

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 Eq. (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 is lower, apart from the last case 10, where the full graph

variant outperformed its relaxed counterpart. We also extracted a MST 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 Fig. 4 show that the best performances of TRE of 1.16 mm were achieved for $v = 0.15$ and $v = 0.1$. Further relaxation of the graph deteriorates TRE for the Dir-Lab dataset. The visualization of the registration results for our method, for the most challenging Dir-Lab case 8, is shown in Fig. 5. Distribution of the landmark error for the same case after registration using our method is presented in Fig. 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. 50, we calculated the gradients of the deformation field with respect to x , y , and z , similarly to calculating the Jacobian of the deformations. Subsequently, we decomposed the deformation field gradients into eigenvalues and found their maximum and minimum principal stretch components, γ_1 and γ_3 ,

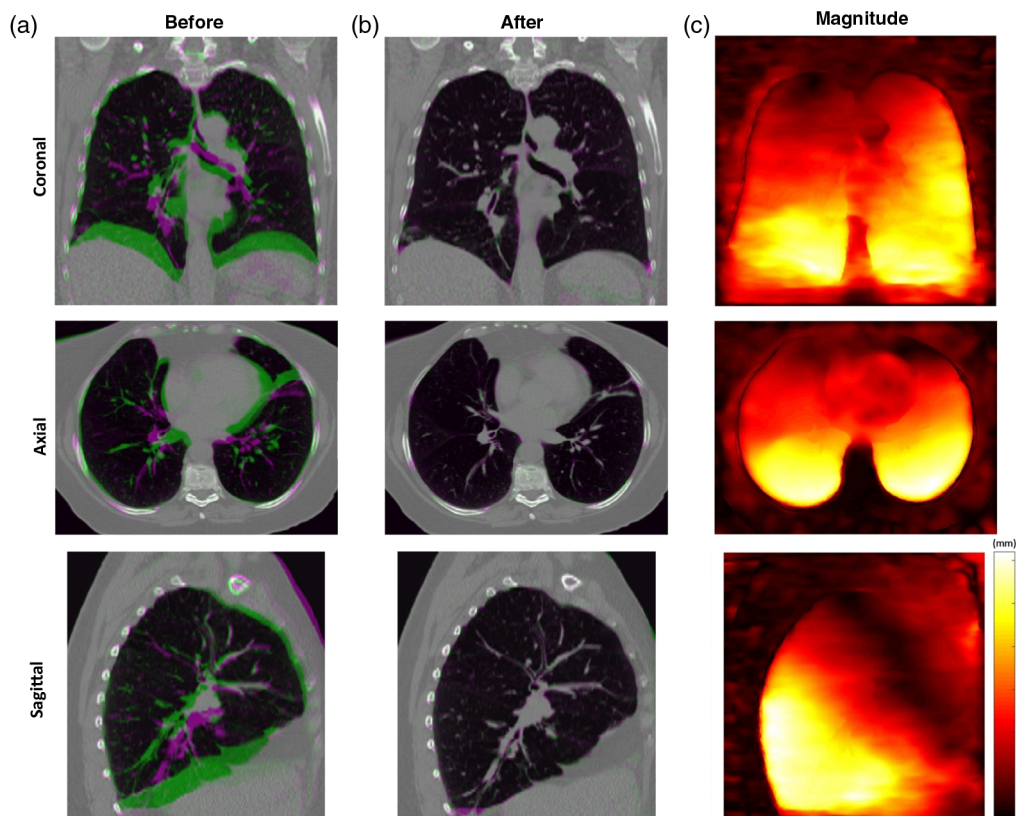


Fig. 5 Registration results for the proposed method for the most challenging Dir-Lab case 8 in axial, coronal, and sagittal view. On the column (a) an overlay of full inhale (green) and full exhale (magenta) before applying the registration is presented. The column (b) shows the same slices after performing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

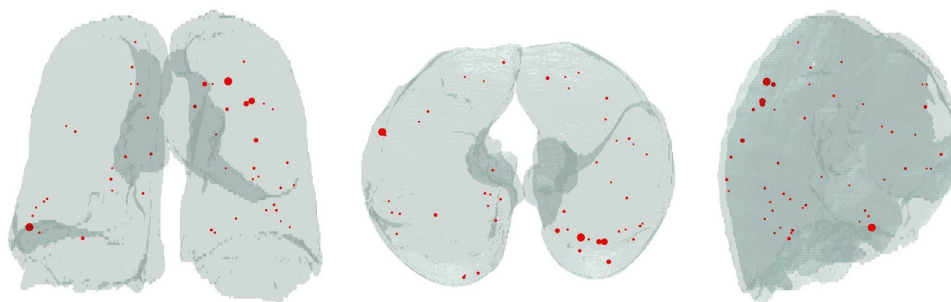


Fig. 6 Visualization of the landmark error distribution for the most challenging case 8 after registration using our method in all three views with the lung mask in the background. Only landmarks with an error higher than 1.5 mm are displayed. The size of the points corresponds to the landmark error.

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 Fig. 7. The highest value of shear stretch $\gamma_{\max} > 5$ almost everywhere at the 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 Fig. 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 with Other Methods

We compared our method against other, most closely related methods in the field, based on the TRE reported for the Dir-Lab dataset. These are continuous optimization-based Demons method enhanced with bilateral-filtering (BLF),¹⁶ dense displacement sampling method using belief propagation on an MST (deeds),¹⁸ and supervoxel-based with belief propagation method applied over an MST (SBP),²⁰ where the

MST 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 1.16 mm, compared with 1.96 mm for BLF, 1.43 mm for deeds, and 1.23 mm for SBP. For all of the cases, the proposed method achieved the lowest TRE, except case 8, where the best result was achieved with the SBP method. The numerical results are presented in Fig. 8. Results achieved for our method show statistically significant improvement compared with 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 deformable image registration method that 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

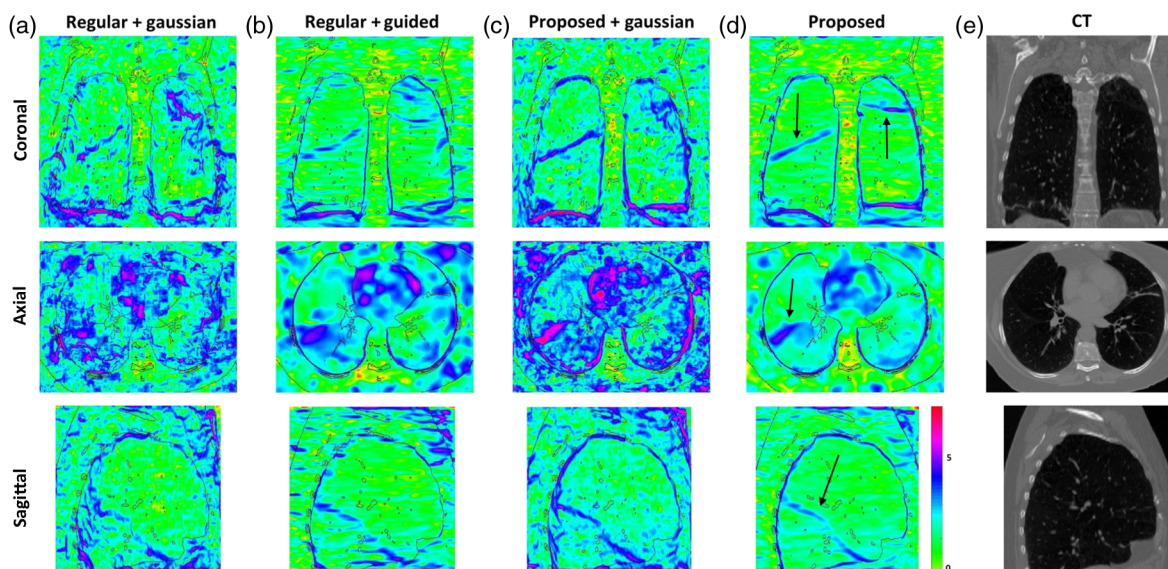


Fig. 7 Example of maximum shear stretch calculated for the most challenging case 8 shown in coronal, axial, and sagittal view. We compared the results for different variants of the proposed method, (a) the RG with Gaussian filtering, (b) the RG with guided image filtering, (c) supervoxels-based image representation with Gaussian filtering and (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.

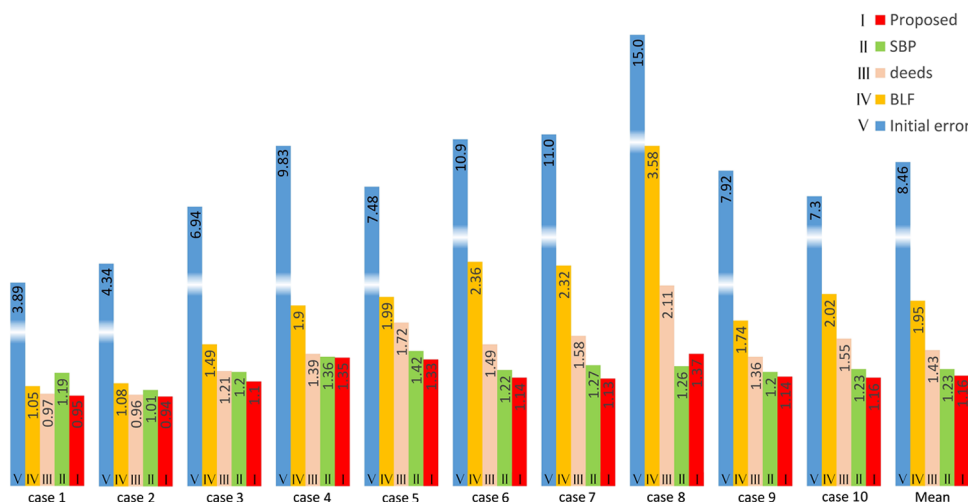


Fig. 8 Comparison of the proposed method with the closest implementations based on the mean TRE for Dir-Lab dataset. For each case, the TRE before registration (initial) and after registration are shown: our proposed method, after BLF combined with Demons (BLF),¹⁶ after the dense displacement sampling method based on belief propagation on an MST (deeds),¹⁸ and after supervoxel-based belief propagation (SBP).²⁰ Mean results show that our proposed method outperforms all of the compared methods on average and in all individual cases apart from case 8.

justification for the chosen approach by first presenting the results for RG 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 supervoxels were imperfect cubes. Nevertheless, it only emphasizes the fact that the perfectly regular cubic representation would further deteriorate the results. For the RG image representation with Gaussian filtering, we achieved TRE of 1.87 mm, 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 1.36 mm, 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 RG. The relative difference between them, compared with the initial TRE of 15 mm, may not seem large; however, visual assessment of the magnitude of the displacement field from Fig. 3 illustrates how much this is advantageous in practice. The displacement field for RG is heavily distorted at lung borders, and the shape of the lungs is not recognizable, 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 RG counterpart. Surprisingly, the supervoxel image representation with Gaussian filtering variant has achieved the same TRE (1.36 mm), as for the RG 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 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 MST-based version were expected, based on the previously reported results in the literature,^{18,20} which positions it well between the RG/Gaussian filtering versions and the proposed method. Coincidentally, the same average TRE (1.23 mm) is achieved in our test as 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 1.19 mm), which were further improved by its relaxed variant (TRE 1.16 mm). 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 Fig. 9, showing the gradual improvement in the results. As an additional experiment, we investigated the influence of the intensity term in the piecewise smoothness term of the energy formulation in Eq. (8). The substitution of the proposed linear intensity term with a quadratic one resulted in TRE of 1.22 mm, compared with 1.16 mm for the linear term. This experiment

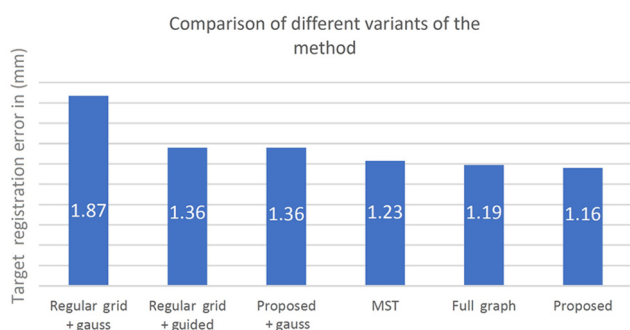


Fig. 9 Comparison of different variants of the method based on TRE in (mm).

supports our proposed formulation of the energy in Eq. (8). In terms of the time, the performance of the proposed method in its current suboptimal and nonparallelized MATLAB implementation varied from 63 to 121 min per pairwise registration, depending on the Dir-Lab case.

In our comparison with other methods, we used the closest methods in approach, while at the same time considered 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 3-D registration to be up to 24 h. In our case, it was between 63 and 121 min, depending on the volume size. It was just slightly over 1 h for images having similar volumes with the ones from Ref. 29. The methods proposed in Refs. 31 and 32 have been presented for 2-D cases only. They could potentially be extended to 3-D 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. 51 with TRE of 0.94 mm 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 the image registration field, where discrete optimization-based methods perform very well and in some cases favorably compared with their continuous optimization-based counterparts. Our proposed framework contributes to the field addressing the aims summarized in Sec. 1.3. The proposed method creates a promising framework for more advanced applications, such as multimodal image fusion, where images originating from different modalities could be combined. In the 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 they could be well distinguishable based on the deformation field, as show in Fig. 7. In our opinion, the method may be applicable to other medical image registration where sliding motion appears, such as multimodal lung registration, liver, or cardiac motion correction. Supervoxel-based image representation appears to be still a very promising method; however, it has not yet been fully explored in image registration.

Acknowledgments

A.S., B.P., J.A.S., and V.G. gratefully acknowledge funding from the CRUK and EPSRC Cancer Imaging Centre in Oxford (CICO), C5255/A16466. A.S. would like to acknowledge Professor M. Pawan Kumar and Professor Olga Veksler for their advice and suggestions, and George Qian for proofreading the manuscript. 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. The authors have no relevant financial interests in this article and no potential conflicts of interest to disclose.

References

1. B. Fulkerson, A. Vedaldi, and S. Soatto, "Class segmentation and object localization with superpixel neighborhoods," in *IEEE 12th Int. Conf. on Computer Vision*, pp. 670–677 (2009).
2. H. Kim 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," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2010)*, pp. 1253–1260 (2010).
5. J. A. Schnabel 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 et al., "Volume adjustment of lung density by computed tomography scans in patients with emphysema," *Acta Radiol.* **45**(4), 417–423 (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. Imaging* **32**(11), 2114–2126 (2013).
8. R. M. Kaus et al., "Estimation of organ motion from 4D CT for 4D radiation therapy planning of lung cancer," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2004)*, Vol. 3217, pp. 1017–1024 (2004).
9. E. Weiss et al., "Tumor and normal tissue motion in the thorax during respiration: analysis of volumetric and positional variations using 4D CT," *Int. J. Radiat. Oncol. Biol. Phys.* **67**, 296–307 (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. Imaging* **31**(2), 153–163 (2012).
12. B. Horn and B. Schunck, "Determining optical flow," *Artif. Intell.* **17**, 185–203 (1981).
13. A. Schmidt-Richberg et al., "Estimation of slipping organ motion by registration with direction-dependent regularization," *Med. Image Anal.* **16**, 150–159 (2012).
14. H. Y. Baluwala et al., "Toward physiologically motivated registration of diagnostic CT and PET/CT of lung volumes," *Med. Phys.* **40**(2), 021903 (2013).
15. L. Risser 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. Papiez et al., "An implicit sliding-motion preserving regularisation via bilateral filtering for deformable image registration," *Med. Image Anal.* **18**, 1299–1311 (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 et al., "MRF-based deformable registration and ventilation estimation of lung CT," *IEEE Trans. Med. Imaging* **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," in *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015)*, Vol. 9350, pp. 338–345 (2015).
20. M. P. Heinrich 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. Vision* **70**(2), 109–131 (2006).
22. P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vision* **59**(2), 167–181 (2004).
23. B. Glocker 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," *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2011)*, pp. 557–565 (2011).
25. A. Raj and R. Zabih, "A graph cut algorithm for generalized image deconvolution," in *IEEE Int. Conf. on Computer Vision (ICCV 2005)*, Vol. 2, pp. 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," in *Proc. Ninth IEEE Int. Conf. on Computer Vision*, Vol. 2, pp. 1033–1040 (2003).

28. T. W. H. Tang and A. C. S. Chung, "Non-rigid image registration using graph-cuts," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2007)*, pp. 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. Chriet, "Landmark-based non-rigid registration via graph cuts," *Lect. Notes Comput. Sci.* **4633**, 166–175 (2007).
31. R. W. K. So and A. C. S. Chung, "Non-rigid image registration by using graph-cuts with mutual information," in *17th IEEE Int. Conf. on Image Processing (ICIP 2010)*, pp. 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," in *18th IEEE Int. Conf. on Image Processing (ICIP 2011)*, pp. 717–720 (2011).
33. R. Achanta et al., "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(11), 2274–2282 (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 et al., "Graph cuts-based registration revisited: a novel approach for lung image registration using supervoxels and image-guided filtering," in *IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW 2016)*, pp. 592–599 (2016).
37. A. Vedaldi and S. Soatto, "Quick shift and kernel methods for mode seeking," *Lect. Notes Comput. Sci.* **5305**(4), 705–718 (2008).
38. V. Machairas et al., "Waterpixels," *IEEE Trans. Image Process.* **24**(11), 3707–3716 (2015).
39. M. P. Heinrich et al., "MIND: modality independent neighbourhood descriptor for multi-modal deformable registration," *Med. Image Anal.* **16**, 1423–1435 (2012).
40. J. Ruhaak et al., "Highly accurate fast lung CT registration," *Proc. SPIE* **8669**, 86690Y (2013).
41. J. Vandemeulebroucke et al., "Automated segmentation of a motion mask to preserve sliding motion in deformable registration of thoracic CT," *Med. Phys.* **39**(2), 1006–1015 (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 et al., "Nonrigid registration using free-form deformations: application to breast MR images," *IEEE Trans. Med. Imaging* **18**, 712–721 (1999).
44. C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Sixth Int. Conf. on Computer Vision*, pp. 839–846 (1998).
45. B. W. Papież et al., "Liver motion estimation via locally adaptive over-segmentation regularization," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015)*, Vol. 9351, pp. 427–434 (2015).
46. M. Chen et al., "A simple fixed-point approach to invert a deformation field," *Med. Phys.* **35**(1), 81–88 (2008).
47. R. Castillo et al., "A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets," *Phys. Med. Biol.* **54**(7), 1849–1870 (2009).
48. E. Castillo et al., "Four-dimensional deformable image registration using trajectory modeling," *Phys. Med. Biol.* **55**(1), 305–327 (2009).
49. J. Ashburner, "A fast diffeomorphic image registration algorithm," *NeuroImage* **38**(1), 95–113 (2007).
50. R. E. Amelon et al., "A measure for characterizing sliding on lung boundaries," *Ann. Biomed. Eng.* **42**(3), 642–650 (2014).
51. L. Knig and J. Ruhaak, "A fast and accurate parallel algorithm for non-linear image registration using normalized gradient fields," in *IEEE 11th Int. Symp. on Biomedical Imaging (ISBI 2014)* (2014).

Adam Szmul is a DPhil candidate in biomedical engineering at the University of Oxford, working at the Institute of Biomedical Engineering. He is mainly focused on medical image registration and motion correction, particularly in lung cancer applications. His work involves single-modality applications, like CT to CT image registration, as well as multimodal image fusion, including a range of different structural and functional imaging modalities.

Bartłomiej W. Papież: Biography is not available.

Andre Hallack received his DPhil degree in healthcare innovation at the University of Oxford, where he conducted research on nonlinear motion correction of dynamic radiological cancer images, in special MRI and US sequences. His focus is on model-based similarity measure, as well as, dense feature transforms.

Vicente Grau received his PhD from the Universidad Politecnica de Valencia, Spain. After 2 years as a postdoctoral researcher at Brigham and Women's Hospital and LSU Health Sciences Centre, he joined the University of Oxford, where he is now a professor of biomedical image analysis within the Institute of Biomedical Engineering. His research focuses on the development of image analysis algorithms for disease characterization, in particular for pulmonary and cardiac applications.

Julia A. Schnabel holds a chair in computational imaging in the Department of Biomedical Engineering, King's College London. From 2007, she was an associate professor and from 2015 a professor in the Department of Engineering Science, University of Oxford. Her research interests include medical imaging, computer vision, and machine learning. She is an associate editor of the *IEEE Transactions on Medical Imaging*, the *IEEE Transactions on Biomedical Engineering*, and editorial board member of *Medical Image Analysis*.