



HHS Public Access

Author manuscript

IEEE Trans Med Imaging. Author manuscript; available in PMC 2017 August 08.

Published in final edited form as:

IEEE Trans Med Imaging. 2017 August ; 36(8): 1650–1663. doi:10.1109/TMI.2017.2688377.

Pulmonary Lobe Segmentation with Probabilistic Segmentation of the Fissures and a Groupwise Fissure Prior

Felix J.S. Bragman,

Centre for Medical Image Computing, Department of Medical Physics & Bioengineering,
University College London, WC1 6BT London, United Kingdom

Jamie R. McClelland,

Centre for Medical Image Computing, Department of Medical Physics & Bioengineering,
University College London, WC1 6BT London, United Kingdom

Joseph Jacob,

Centre for Medical Image Computing, Department of Medical Physics & Bioengineering,
University College London, WC1 6BT London, United Kingdom

Department of Radiology in the Mayo Foundation for Medical Education and Research, Mayo
Clinic, Rochester, Minnesota, USA

John R. Hurst, and

UCL Respiratory, University College London, WC1 6BT London, United Kingdom

David J. Hawkes

Centre for Medical Image Computing, Department of Medical Physics & Bioengineering,
University College London, WC1 6BT London, United Kingdom

Abstract

A fully automated, unsupervised lobe segmentation algorithm is presented based on a probabilistic segmentation of the fissures and the simultaneous construction of a population model of the fissures. A two-class probabilistic segmentation segments the lung into candidate fissure voxels and the surrounding parenchyma. This was combined with anatomical information and a groupwise fissure prior to drive non-parametric surface fitting to obtain the final segmentation.

The performance of our fissure segmentation was validated on 30 patients from the COPDGene cohort, achieving a high median F_1 -score of 0.90 and showed general insensitivity to filter parameters. We evaluated our lobe segmentation algorithm on the LOLA11 dataset, which contains 55 cases at varying levels of pathology. We achieved the highest score of 0.884 of the automated algorithms. Our method was further tested quantitatively and qualitatively on 80 patients from the COPDGene study at varying levels of functional impairment. Accurate segmentation of the lobes is shown at various degrees of fissure incompleteness for 96% of all cases. We also show the utility of including a groupwise prior in segmenting the lobes in regions of grossly incomplete fissures.

Keywords

Lobe segmentation; fissure segmentation; pulmonary image analysis

I. Introduction

Segmentation of the pulmonary lobes can facilitate the localisation and quantification of respiratory diseases and is of particular interest in Chronic Obstructive Pulmonary Disease (COPD). COPD can alter the structure of the lung through emphysematous destruction of lung parenchyma. The speed with which local pulmonary damage evolves can vary between patients with COPD [1], [2], yet the heterogeneity of local disease progression [2] may not be captured in lung physiologic indices that quantify function at a global level [1]. CT-based lobe segmentation provides an anatomically consistent reference frame for the quantitative analysis of parenchymal damage across large cohorts of patients and negates the requirement of a groupwise space for analysis. Knowledge of the underlying lobar distribution of disease may allow the identification of subtle COPD phenotypes or help identify patients that would benefit from interventions such as lung volume reduction surgery [3].

To characterise CT disease extent on a lobar basis, it is necessary to identify the pulmonary fissures. The fissures consist of invaginations of visceral pleura, which extend from the lung periphery to the lung hilum and separate the right and left lung into five lobes. The oblique and horizontal fissures divide the right lung into three lobes (upper, middle and lower) whilst the left oblique fissure divides the left lung into upper and lower lobes. When visible on CT, the fissures appear as bright, solid lines. They represent two apposed layers of visceral pleura, which are usually devoid of airways and vascular structures [4]. However, the appearances of the fissures on CT can be variable in the general population [5]. A developmental failure of pleural invaginations can result in congenitally absent or incomplete fissures [4] (Fig. 1b). Various pathological processes may damage the pleural surfaces disrupting the integrity of the fissures [4] (Fig. 1c).

Automated lobar segmentation is most reliable when fissures are complete [6]. In cases with incomplete fissures, various methods have been developed that draw information from pulmonary anatomy and atlases. Lobe segmentation algorithms can be broadly categorised as either supervised [7]–[12] or unsupervised [13]–[16]. In our study, we extend the definition of supervised methods to encompass any algorithm that requires prior manual labelling to determine optimal fissure properties or the construction of anatomical atlases. Segmentation algorithms can be further subdivided on the basis of the segmentation of auxiliary structures. Methods can be dependent [7], [10], [13]–[17] or independent [12], [18] of the information provided by the airway and vascular trees. Algorithms can also be classified based on their dependence on anatomical atlases [10]–[13], [18] or whether the method is uniquely performed in the patient-space [7], [14].

Fissure segmentation can be classified as supervised when posed as a classification task or unsupervised when applied with a filter. A major shortcoming of filters is their reliance on arbitrary thresholds for segmentation. The inclusion of fixed segmentation thresholds may ignore potential fissure voxels or include excessive false positive voxels. Such thresholds are

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

often compromised when lung attenuation values themselves are variably influenced by a range of factors including the CT reconstruction algorithm, CT slice thickness and patient inspiratory effort. The issue of the removal of false-positive voxels and its dependence on prior knowledge is a further limitation of several algorithms [7], [12], [19], [20]. The likelihood function of Lassen et al. [7] requires prior knowledge of fissure Hessian eigenvalues and may lead to an over-segmentation of fissure voxels that cannot be corrected through post-processing techniques. The work of Wiemker et al. [19] requires knowledge of the underlying Hounsfield intensity distribution of the fissures with no data-driven method presented to determine these parameters. Ross et al. [12] sample the image domain to detect the most likely fissure surface based on Hessian eigenvalues and a maximum a posteriori estimation. Their technique requires a lobe boundary shape model based on manually segmented data to improve their fissure discrimination. The method is similar to the formulation of van Rikxoort et al. [21], which requires prior knowledge of manually labelled voxels to build a classifier. Manual annotation of data is time consuming and impractical in routine clinical practice. Moreover, it does not follow that a training dataset built on a single set of scans will generalise to a new cohort derived using different scanners, with varying reconstruction kernels. Such a constraint is also apparent in the likelihood function of Lassen et al. [7]. Our technique however learns the necessary model parameters from the volume being segmented, permitting the development of a robust segmentation tool, applicable across a broad range of datasets.

The use of prior knowledge derived from population models has increased in popularity [12], [18], [22], [23]. Zhang et al. [23] perform lobe segmentation using a single atlas search initialisation. The average fissure surface from a training set is exploited in a fuzzy reasoning system to segment the fissures and the lobes. An alternative multi-atlas selection mechanism has been proposed by van Rikxoort et al. [10]. This selects the most similar atlas to the patient by comparing the patient fissure segmentation to the atlas and exploits a transformation to combine atlas lobe labels with an approximate lobe segmentation. Ross et al. [12] exploit a deformable model in fissure surface extraction. The ability to exploit prior knowledge is an implicit advantage of atlases. However, if the training data is not large enough, this may not correctly model the shape variation within the population. These methods described all require complete segmentations prior to model building, which is a laborious task. We aim to build a simple population model of the fissures negating the need for prior manual labelling without requiring complete fissure segmentations.

The limitations associated with the dependence on manually segmented data, either to train classifiers or build atlases was a major motivation of the work presented. When considering fissure segmentation, there is sufficient data within a single scan to detect the fissures when visible whilst rejecting most false-positives. In view of large-scale studies such as COPDGene [24], CT scans can also be pooled together to produce a prior, which negates the need for complete manual segmentations.

II. Method

We present an automatic lobe segmentation algorithm (Fig.2) based on a probabilistic segmentation of the fissures (Section II-B) and the construction of a groupwise fissure model

(Section II-C). Our study aimed to construct a fissure model (Section II-C) using complete and incomplete fissures to generate a confidence region based on a population. In the context of routine clinical care, new scans can be iteratively added if necessary to help strengthen the population model.

The main technical contributions of this paper are: a) unsupervised probabilistic segmentation of the fissures with iterative false-positive removal, b) the simultaneous construction of a groupwise prior without need for complete manual segmentations and c) post-processing of the airway segmentation to correct errors in seed labelling. An overview of the segmentation is shown in Fig. 2. The lungs, vessel and airway tree are first segmented. This is followed by a segmentation of the fissures using auxiliary tree structures as anatomical priors in a probabilistic setting. The segmented fissures are then combined using a groupwise registration framework to produce a population prior. The anatomical information, the segmented fissure and the groupwise fissure prior are then combined as a cost image for a watershed segmentation.

A. Data pre-processing

Lung masks are obtained with the algorithm of Hu et al. [25]. The vasculature is segmented by considering multiscale vessel filtering [26]. The airways are segmented using region growing via evolution of a wavefront, which iteratively corrects for leakage across the airway wall [14]. It is assumed that the remaining structures after segmentation are the fissures and the parenchyma. A skeletonisation of the airways reveals the branching structure, used to label the lobar bronchi to generate surface fitting seeds. All pre-processing is performed using the Pulmonary Toolkit¹ with standard parameter settings.

B. Probabilistic fissure segmentation

We propose an unsupervised fissure segmentation framework that does not require any training data to classify fissure-voxels whilst negating the need to empirically determine algorithm parameters. We present a simple fissure enhancement filter that does not require any manual observations to set the parameters. We then construct a probabilistic framework to segment the fissures based on this enhancement filter. We assume a generative model between the observed filter result and the underlying segmentation of the lung and that these hidden segmentations exhibit separate Gaussian distributions. The proposed model assumes that the filtered image is generated by a two-class Gaussian mixture model (GMM), where the fissures and the parenchyma are the hidden segmentations that have generated the observed enhancement filter. Parameters of the GMM and the underlying segmentation are determined through application of the Expectation-Maximisation (EM) algorithm.

1) Multi-scale fissure enhancement filter—This filter aims to distinguish fissure-like voxels from surrounding structures. If we consider an image volume \mathcal{Q}_t from the set of images $\mathcal{Q} = \{\mathcal{Q}_1, \mathcal{Q}_2, \dots, \mathcal{Q}_T\}$, the Hessian matrix (\mathbf{H}_t) at a scale σ is obtained by considering the second derivative of \mathcal{Q}_t convolved with a Gaussian kernel $G(\sigma)$.

¹<https://github.com/tomdoel/pulmonarytoolkit>

The width of the pleural cavity is likely to vary. The filter is embedded in a multi-scale framework to capture this variation. A voxel-wise eigen-analysis of $\mathbf{H}_t(\mathcal{Q}_t; \sigma)$ yields scale-dependent eigenvalues λ_1, λ_2 and λ_3 , which are, ordered based on their magnitude such that $|\lambda_1| \geq |\lambda_2| \geq |\lambda_3|$, with respective eigenvectors $\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2$ and $\hat{\mathbf{u}}_3$. We adapt the vessellness filter of Frangi et al. [26] to capture voxels exhibiting a fissure-like shape.

As a sheet-like structure, a candidate fissure voxel will ideally be represented by a very large $|\lambda_3| \gg 0$ with $\lambda_2 \approx \lambda_1 \approx 0$. We aim to enhance voxels with this relationship whilst suppressing other auxiliary structures using the following parameters:

$$\mathcal{R}_a = \frac{|\lambda_1|}{\sqrt{|\lambda_2 \lambda_3|}}, \mathcal{R}_b = \frac{|\lambda_2|}{|\lambda_3|} \quad (1)$$

The parameter \mathcal{R}_a helps differentiate spherical structures from plate-like and tubular structures. To differentiate plates from tubes, \mathcal{R}_a is combined with the parameter \mathcal{R}_b , which seeks to enhance structures exhibiting a plate-like aspect ratio. A scale dependent filter (Eq. 2) is obtained by combining both terms:

$$\mathcal{F}_t(\sigma) = \mathbb{I}(\lambda_3(\sigma)) \cdot \left\{ \exp\left(-\frac{\mathcal{R}_a(\sigma)^2}{A}\right) \exp\left(-\frac{\mathcal{R}_b(\sigma)^2}{B}\right) \right\} \quad (2)$$

where $\mathbb{I}(\lambda_3(\sigma))$ is an indicator function such that $\mathbb{I}(\lambda_3(\sigma)) = 0$ when $\lambda_3(\sigma) > 0$ to seek only bright features. The parameters A and B control the sensitivity of each parameter $\mathcal{R}_{a,b}$. Small values of A and B (≈ 0) will only be sensitive to voxels with ideal values for the filter parameters ($\mathcal{R}_{a,b} \rightarrow 0$). Larger values for A and B ($\rightarrow 0.5$) will enhance voxels with less ideal relationships with the caveat of enhancing more false positives (Section IV-A). A final value of \mathcal{F}_t is found by computing the scale σ which maximises \mathcal{F}_t at a voxel \mathbf{x} :

$$\mathcal{F}_t(\mathbf{x}) = \max_{\sigma_{\min} \leq \sigma \leq \sigma_{\max}} \mathcal{F}(\mathbf{x}; \sigma). \quad (3)$$

There is a deviation from the ideal plate-like eigenvalue relationship in the presence of partial-volume effects, image noise and patient motion, which may result in a loss in the discriminating power of \mathcal{F}_t . We assume that both tissue classes are hidden segmentations that generate a range of values stemming from Gaussian distributions. We aim to capture these class distributions to accurately segment the fissures, even when the filter response is poor at a fissure voxel by considering local neighbourhood properties.

2) Fissure segmentation using a Gaussian Mixture Model—The output of the filter \mathcal{F}_t is parameterised by a two-class GMM. The two underlying distributions in the signal correspond to the fissures and all other remaining structures. We assume that the fissures and lung tissue are hidden segmentations (z) that give rise to the observed values y of \mathcal{F}_t . The segmentation can be modelled as a random process with a probability density function $f(z | \Phi_z)$ with parameters Φ_z . The total filter signal has a probability density function $f(y | z, \Phi_y)$

parameterised by the model parameters Φ_y . The goal is to estimate the segmentation z by the parameters $\Phi = \{\Phi_y, \Phi_z\}$. This is performed using the EM algorithm by estimating the maximum-likelihood parameters $\hat{\Phi}$ via maximisation of the log-likelihood

$$\hat{\Phi} = \arg \max_{\Phi} \log f(y|\Phi) \quad (4)$$

We consider the image model of Van Leemput et al. [27]. The index of a voxel x is $i \in \{1, 2, \dots, n\}$ where n is the number of voxels within the lung mask. There are $K = 2$ classes (fissure and tissue). The class of the i^{th} voxel is defined as $z_i = e_k$. The variable e_k represents the class membership e.g. $e_{k=1}$ defines the fissure class and $e_{k=2}$ is the surrounding lung parenchyma. The response of the filter at voxel x_i is y_i . The filter values belonging to each class k are assumed to be normally distributed after log transformation with mean μ_k and standard deviation σ_k such that $\phi_k = \{\mu_k, \sigma_k\}$. The vector $\Phi_y = \{\phi_{k=1}, \phi_{k=2}\}$ represents the model parameters for both tissue classes. The overall probability density for y_i is defined as a mixture of normal distributions,

$$f(y_i|\Phi_y) = \sum_k G_{\sigma_k}(y_i - \mu_k) f(z_i = e_k) \quad (5)$$

where G_{σ_k} represents the k^{th} class zero-mean normal distribution with standard deviation σ_k and $f(z_i = e_k)$ is the class prior probability of a voxel x_i . By assuming statistical independence over all voxels $x \in \mathcal{Q}$, the overall joint probability density is given by

$$f(y|\Phi_y) = \prod_i f(y_i|\phi_y) \quad (6)$$

The maximum-likelihood estimates for Φ_y are found using Eq. 4 by seeking the parameters that maximise Eq. 6, giving the following update equations for the model parameters

$$\mu_k^{(m+1)} = \frac{\sum_i p_{ik}^{(m+1)} y_i}{\sum_i p_{ik}^{(m+1)}} \quad (7)$$

$$\left(\sigma_k^{(m+1)}\right)^2 = \frac{\sum_i p_{ik}^{(m+1)} \left(y_i - \mu_k^{(m+1)}\right)^2}{\sum_i p_{ik}^{(m+1)}} \quad (8)$$

where

$$p_{ik}^{(m+1)} = \frac{f(y_i|z_i=e_k, \Phi_y^{(m)}) f(z_i=e_k)}{\sum_{j=1}^K f(y_i|z_i=e_j, \Phi_y^{(m)}) f(z_i=e_j)} \quad (9)$$

is a probabilistic estimation of the hidden data z_i of class k at voxel x_i given the filter value y_i . The class k is iterated with the class index j and m denotes the EM iteration number.

The segmentation resulting from Eq. 9 can be sensitive to noise, image artefacts and false positives as the segmentation is only based on y . Priors ($\pi_{t,ik}$) that incorporate probabilistic information about the segmentation are typically added to the model [27], [28]. In the context of this work, we can quantify the likelihood of the fissure location (Fig. 4b) based on the vessel and airway tree. We quantify the vessel density (v_d) [14], which is obtained by applying a strong (10mm isotropic) Gaussian filter to the vesselness filter of Frangi et al. [26]. This measure is inverted and scaled in the range [0, 1] using min-max scaling such that regions of low vessel density are close to 1. Airway density (a_d) is estimated by computing the Euclidean distance transform to the airway segmentation and is normalised using min-max scaling to the range [0, 1] such that regions of high distance to the airways are close to

1. The fissure likelihood measure is defined as $\pi_{t,ik=1} = (a_d + v_d)/2$ and the tissue likelihood is $\pi_{t,ik=2} = 1 - \pi_{t,ik=1}$. The subscript t is dropped in Eq. 10 and 11 for convenience. The anatomical information is integrated into Eq. 5 by setting $f(z_i=e_k) = \pi_{ik}$.

Information about lung structure, spatial smoothness and morphology can be also be enforced by considering a Markov Random Field (MRF) regularisation term (U_{MRF}). The probability of a voxel i belonging to tissue class k is now dependent on the first-order

neighbours \mathcal{N}_i . The neighbourhood system at a voxel x_i is defined as $\mathcal{N}_i = \{\mathcal{N}_i^x, \mathcal{N}_i^y, \mathcal{N}_i^z\}$ in the face-connected neighbourhood. The likelihood term (π) is now augmented with an MRF that is dependent on the probability and curvature of neighbouring voxels. By employing the formulation of Van Leemput et al. [27], Eq. 9 is updated to

$$p_{ik}^{(m+1)} = \frac{f(y_i|z_i=e_k, \Phi_y^{(m)}) f(z_i=e_k|p_{\mathcal{N}_i}^{(m)}, \Phi_z^{(m)})}{\sum_{j=1}^K f(y_i|z_i=e_j, \Phi_y^{(m)}) f(z_i=e_j|p_{\mathcal{N}_i}^{(m)}, \Phi_z^{(m)})} \quad (10)$$

with

$$f(z_i=e_k|p_{\mathcal{N}_i}^{(m)}, \Phi_z^{(m)}) = \frac{\pi_{ik} e^{-\beta_i U_{\text{MRF}}(e_k|p_{\mathcal{N}_i}^{(m)}, \Phi_z^{(m)})}}{\sum_{j=1}^K \pi_{ij} e^{-\beta_i U_{\text{MRF}}(e_j|p_{\mathcal{N}_i}^{(m)}, \Phi_z^{(m)})}}. \quad (11)$$

where the MRF term $U_{\text{MRF}}(z_i|p_{\mathcal{N}_i}, \Phi_z)$ is an energy function dependent on $\Phi_z = \{G\}$ and the MRF weight β_i is kept constant for all voxels i .

Structure in the segmentation is enforced by considering neighbourhood probabilities ($p_{\mathcal{N}}$) and constraints on the curvature of the surface (f_{dot_k}) (Fig. 3b). The fissure surface exhibits low local curvature, which will be captured in the eigenvector $\hat{\mathbf{u}}_3$. A weight in the MRF energy term is introduced based on the dot product of the eigenvectors of a neighbouring voxel x_l with the center voxel x_i such that

$$f_{dot}(x_l, x_i)_k = \begin{cases} 1 - \exp\left(-\frac{(|\hat{\mathbf{u}}_{x_l,3} \cdot \hat{\mathbf{u}}_{x_i,3}|)^6}{0.25}\right) & \text{if } k=1 \\ \exp\left(-\frac{(|\hat{\mathbf{u}}_{x_l,3} \cdot \hat{\mathbf{u}}_{x_i,3}|)^6}{0.25}\right) & \text{if } k=2 \end{cases} \quad (12)$$

When considering the fissure class ($k = 1$), the weight will tend to 0 as dissimilarity in the local curvature increases ($1 - \exp(-|\hat{\mathbf{u}}_{l,3} \cdot \hat{\mathbf{u}}_{i,3}|) \rightarrow 0$). If neighbouring voxels x_l and x_i have similar curvature, $f_{dot}(x_l, x_i)_{k=1}$ will tend to 1. The goal of this function is to force candidate fissure voxels to have approximately equal curvature whilst negatively weighting the probabilities of false-positives with non-equal local curvature.

The possibility of anisotropic voxel sizes and slice spacing is considered with the connection-strength factor (s) introduced by Cardoso et al. [28], defined as

$s = \{s_x, s_y, s_z\} = \{1/d_x, 1/d_y, 1/d_z\}$ based on real-world distances between the centre of neighbouring voxels. Closer voxels will yield higher weights in the MRF. The total energy (U_{MRF}) (Fig. 3) for a face-connected neighbourhood \mathcal{N}_i centered at voxel x_i in Eq. 11 is defined as

$$U_{MRF} \left(e_k | p_{\mathcal{N}_i}^{(m)}, \Phi_z^{(m)} \right) = \sum_{j=1}^K \mathbf{G}_{kj} \left(\sum_{l \in \mathcal{N}_i^x} s_x \cdot f_{dot}(l, i)_j \cdot p_{l,j} + \sum_{l \in \mathcal{N}_i^y} s_y \cdot f_{dot}(l, i)_j \cdot p_{l,j} + \sum_{l \in \mathcal{N}_i^z} s_z \cdot f_{dot}(l, i)_j \cdot p_{l,j} \right) \quad (13)$$

where \mathbf{G} represents a K by K matrix whose elements G_{kj} represent the transition energy between tissue classes k and j and the the subscript l denotes the neighbourhood iterator in each direction x, y, z . Since this is a two-class problem, the matrix \mathbf{G} is set up with diagonal elements equal to 0, off-diagonal elements set to 1 and is a constant in our framework.

Initial parameters for the mixture model (Eq. 7 and 8) are set to $\mu^{(m=0)} = \{0.10, 0.90\}$ on the assumption that both class distributions are significantly different. The class standard deviations are initialised as the original standard deviation of the image filter ($\sigma^{2(m=0)} = \{\sigma^2(\mathcal{F}_0), \sigma^2(\mathcal{F}_0)\}$). Initial values for the MRF energy weights (Eq. 11) are set to an even split of 0.5. A termination criteria based on the ratio of likelihood change is set to $\epsilon = 10^{-3}$.

The parameter β in Eq. 11 controls the regularisation strength. To mitigate dependence of the segmentation on a user-defined choice, the segmentation is performed iteratively whilst increasing the strength of β . The percentage of high probability fissure voxels ($p_{i,k=1} > 0.75$)

with respect to the number of voxels n is quantified. This percentage decreases as β rises leading to a fall in false-positive fissure voxels. The initial regularisation is $\beta = 0.75$. The EM framework is run and the percentage high probability voxels is quantified. If there is a convergence of this percentage, the segmentation framework terminates. Otherwise, the regularisation is automatically increased by 0.50. Convergence is defined when the percentage has not fallen by at least 2% in 5 successive iterations. The output of the framework after convergence yields probabilistic fissure ($\mathcal{P}_{t,\text{fissure}}$) and tissue ($\mathcal{P}_{t,\text{tissue}}$) maps for each image \mathcal{Q}_t . To obtain a binary segmentation of the fissures (\mathcal{S}_t), a two-pass analysis is performed. A connected component analysis of $\mathcal{P}_{t,\text{fissure}}$ using a face-connected neighbourhood is performed. Firstly, all components with a median probability below 0.50 are removed then all components below a volumetric threshold of 0.50mL are discarded.

C. Groupwise fissure prior

In a given patient cohort, there will be a range of cases with incomplete fissures, which will complicate the segmentation of the lobes. The goal is to combine all segmented fissures into a groupwise space to create an average fissure model to help guide the lobe segmentation in problematic cases.

1) Groupwise registration—Given a set of T patients $\mathcal{Q} = \{\mathcal{Q}_1, \mathcal{Q}_2, \dots, \mathcal{Q}_T\}$, a common average space $\Omega_{\hat{\mathcal{Q}}}$ is computed by iteratively registering the set of patients \mathcal{Q} to the Fréchet mean. The output is an average image $\hat{\mathcal{Q}}_{\text{avg}}$ and a set of forward and backward transformations ($\varphi_{\mathcal{Q}_T, \hat{\mathcal{Q}}}$ and $\varphi_{\hat{\mathcal{Q}}, \mathcal{Q}_T}$) such that $\varphi_{\mathcal{Q}_T, \hat{\mathcal{Q}}} : \Omega_{\mathcal{Q}_T} \rightarrow \Omega_{\hat{\mathcal{Q}}}$ and $\varphi_{\hat{\mathcal{Q}}, \mathcal{Q}_T} : \Omega_{\hat{\mathcal{Q}}} \rightarrow \Omega_{\mathcal{Q}_T}$.

The registration is performed using the NiftyReg software package [29]. The algorithm is motivated by the work of Ashburner et al. [30]. All patients in \mathcal{Q} are initially registered to an initial template image $\hat{\mathcal{Q}}_{\text{avg}}^{(n=0)}$, which is chosen at random from the set of images \mathcal{Q} . The initial average space is created using a rigid registration. This prevents the atlas from being biased by the geometry of the initial template image. A set of n_1 affine registrations using symmetric block-matching [31] are then performed followed by a set of n_2 non-rigid registrations. The non-rigid registration uses a stationary velocity field, which is parameterised by a cubic B-spline with a 12mm spacing. The locally normalised cross-correlation is used as the similarity with a Gaussian kernel of 50mm. The number of iterations was determined in a pilot experiment by computing the sum of squared differences similarity between successive average images $\hat{\mathcal{Q}}_{\text{avg}}^{(n)}$ and $\hat{\mathcal{Q}}_{\text{avg}}^{(n+1)}$ at iterations n and $n+1$. Convergence of similarity measures occurred after $n_1 = 5$ and $n_2 = 5$ affine and nonrigid registrations.

At each iteration (n) of the algorithm, all patients are registered to the average image $\hat{\mathcal{Q}}_{\text{avg}}^{(n-1)}$.

In order to create the new average image $\hat{\mathcal{Q}}_{\text{avg}}^{(n)}$ and the space $\Omega_{\hat{\mathcal{Q}}^n}$, the inverse average transformation from all patients is computed in the log-Euclidean space. Each patient transformation is demeaned using the inverse average transformation. All patients are subsequently resampled using the demeaned transformations to create $\hat{\mathcal{Q}}_{\text{avg}}^{(n)}$. This new average image $\hat{\mathcal{Q}}_{\text{avg}}^{(n)}$ is obtained by averaging all the resampled images.

2) Construction of the population prior—The set of patients $\mathcal{Q} = \{\mathcal{Q}_1, \mathcal{Q}_2, \dots, \mathcal{Q}_T\}$ will have a corresponding set of binary fissure volumes $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_T\}$. The transformations resulting from the groupwise registration ($\phi_{\mathcal{Q}_t, \hat{\mathcal{Q}}}$ and $\phi_{\hat{\mathcal{Q}}, \mathcal{Q}_t}$) are exploited to build the fissure prior ($\hat{\mathcal{S}}_{avg}$). Each fissure segmentation \mathcal{S}_t is resampled into the groupwise space ($\Omega_{\hat{\mathcal{Q}}}$) using the respective forward transformation $\phi_{\mathcal{Q}_t, \hat{\mathcal{Q}}}$ to yield $\hat{\mathcal{S}}_t$. All resampled fissures ($\hat{\mathcal{S}} = \{\hat{\mathcal{S}}_1, \hat{\mathcal{S}}_2, \dots, \hat{\mathcal{S}}_T\}$) are averaged in the groupwise space to create the average fissure $\hat{\mathcal{S}}_{avg}$. In order to exploit this information to help segment the lobes, the average fissure is resampled using the backwards transformation $\phi_{\hat{\mathcal{Q}}, \mathcal{Q}_t}$ into each patient space ($\Omega_{\mathcal{Q}_t}$) resulting in Π_t . This prior is normalised to the range [0,1] for each patient and is subsequently smoothed using a Gaussian kernel ($\sigma = 2.5\text{mm}$). This produces a prior in the space of each patient, denoting a region where the fissure is expected.

D. Watershed surface fitting

The segmented fissure, the groupwise fissure prior and anatomical information are combined into a cost image inspired by the formulation of Lassen et al. [7]. They create a cost function by combining information from the vessel and airway tree, the segmented fissure and the voxel intensities. We build on this work by extending the cost function to utilise a fissure groupwise prior derived from the population to be segmented.

1) Watershed cost function—The population prior (Π_t) is first combined with the segmented fissure (\mathcal{S}_t). The aim of this step is to produce an initial cost function, with regions of complete and incomplete fissures accentuated using information from the segmentation \mathcal{S}_t and the population prior Π_t . The inverted Euclidean distance function is applied to \mathcal{S}_t to help deal with minor gaps in the segmentation. It is normalised with min-max scaling to the range [0,1] with a value of 1 at the fissure. Only regions in the distance map ($f_{dist}(\mathcal{S}_t)$) within 2.5mm of the fissure are considered. The distance map and the population prior are averaged and convolved with a small Gaussian kernel ($\sigma = 1.0\text{mm}$) to produce a smooth map in Eq. 14. The magnitude of c_1 will be strongest when $f_{dist}(\mathcal{S}_t)$ and Π_t are in the same anatomical location. When there are large gaps in \mathcal{S}_t due to fissure incompleteness, Π_t will provide a local maxima.

$$c_{t,1} = \left(\frac{f_{dist}(\mathcal{S}_t) + \Pi_t}{2} \right) * G(\sigma). \quad (14)$$

This is then combined with the anatomical information ($\pi_{t,k=1} \forall i \rightarrow \pi_{t,k=1}$) and the binary segmentation (\mathcal{S}_t):

$$\mathcal{C}_t = \left(\frac{c_{t,1} + \pi_{t,k=1} + \mathcal{S}_t}{3} \right) \quad (15)$$

The fissure likelihood based on the vessel and airway tree ($\pi_{t,k=1}$) provides a satisfactory estimate for regions of low and high fissure probability and helps guide the segmentation into regions of low vessel and airway density. In addition to c_1 , the original segmentation

(\mathcal{S}_t) is reintroduced. This is performed to produce a high value in \mathcal{C}_t at the segmented fissure since this is the true location of the lobar border.

2) Lobe seed labelling—We employ the method used by Doel et al. [14] to generate initial lobe seed labels from the segmented airway tree. The seed labels are dilated and allowed to grow according to the vessel density map for a limited amount of iterations using the watershed algorithm.

3) Lobe seed labelling post-processing—The segmentation is dependent on the initial seed labels. The quality of the initial labelling can be affected by segmentation failures and errors in the bronchial labelling. If a lobe seed is incorrectly labelled, the resulting segmentation will be erroneous despite successful segmentation of the fissures. To improve the robustness of the pipeline with respect to the airway tree segmentation, the following errors are accounted for: 1) mislabelled branches and 2) unsegmented branches.

a) Seed label correction - labelling errors: To detect and remove mislabelled branches, the centre of mass (CoM) of each set of label seeds is quantified. For each label set, the intra-label distance of each component to the label CoM is computed. The inter-label distance of the components to all other label CoMs are also quantified. The dilated components are removed iteratively to minimise the amount of deleted seeds. A seed is marked for removal if its inter-label distance is smaller than its intra-label distance. The seed with the smallest inter-component distance of all candidate components is marked for removal. This component is discarded if its removal does not cause the number of components for that label to fall below a threshold ($r_{threshold} = 4$). If this threshold is met, the component with the next smallest distance is considered. Once a component is removed, the above method is repeated until removal is no longer possible. This enforces maximum separability between the seeds and removes all potentially erroneously labelled seeds.

b) Seed label correction - airway segmentation errors: If the airway tree segmentation fails, labelling of the branches will not yield the necessary seeds to segment all lobes. In this instance, the anatomical information ($\pi_{t,k=1}$) and its distribution in non-fissure regions (16) is considered. The threshold is considered by analysing the distribution of the prior ($\pi_{t,k=1}$) at the fissure \mathcal{S}_t

$$\pi_{t, threshold} = \mu(\pi_{t,k=1} \in \mathcal{S}_t) - 2\sigma(\pi_{t,k=1} \in \mathcal{S}_t) \quad (16)$$

The low minima regions of $\pi_{t,k=1}$ are defined as $\pi_{t,k=1} < \pi_{t,threshold}$. These correspond to regions of high vessel and airway branching density. A mask of these regions is created and a connected component analysis is performed to extract the regions of local minima. These regions are analysed based on their position within the lung and are exploited to generate new seed labels should the airway tree segmentation fail.

4) Final lobe segmentation—Segmentation of the lobes (\mathcal{L}_t) is obtained by combining the cost image \mathcal{C}_t and the processed seed labels in a watershed segmentation. The lobar

boundaries are smoothed to deal with minor artefacts in the segmentation. This is performed by normalised convolution with a 4.0mm^3 Gaussian kernel.

III. Data

A. Dataset 1

Dataset 1 was based on the LOBe and Lung Analysis 2011 (LOLA11) challenge [32]. It consists of 55 volumetric chest CT scans originating from a variety of source with a range of scans containing serious pathology and abnormalities. The inplane resolution is between 0.53mm and 0.78mm whilst the slice thickness ranges from 0.3 to 1.5mm. The organisers manually segmented the lobes on 9 coronal slices with two human observers and were instructed only to label when the boundaries were visible. The inter-observer agreement between the lobar borders was $1.50\text{mm} \pm 1.28\text{mm}$.

B. Dataset 2

Dataset 2 was used to quantitatively and qualitatively validate the framework on patients with COPD from the COPDGene study [24]. We created a quantitative ($N_{quant} = 30$) and qualitative set ($N_{qual} = 50$) by randomly selecting patients from the study. Minimum thresholds ($> 10\%$) for the level of emphysema in the inspiration scan and gas trapping in the corresponding expiration scan were set in the qualitative cohort to capture cases with significant pathology. The quantitative set averaged $12.8\% \pm 11.12\%$ emphysema whilst the qualitative set averaged $22.50\% \pm 3.60\%$ emphysema.

Analysed scans stem from GE Medical Systems (Light-Speed 16, Lightspeed VCT), Siemens (Sensation 16, Sensation 64 and Definition) and Philips (Brilliance 64) scanners. Scans with the STANDARD (GE), AS+ B31f and B31f (Siemens), and 64 B (Philips) reconstruction algorithms were analysed. Information about the scanning protocols can be viewed at the COPDGene website². The slice thickness of the scans range from 0.62mm to 1.00mm with in-plane dimensions ranging from 0.52 to 0.90mm.

The quantitative cohort was built by manually tracing the fissures in every fifth sagittal slice using ITK-SNAP [33]. The radiologist was asked to manually trace the fissures using three labels. Label 1 was used when the fissures were visible. Label 2 was employed in cases where extrapolation was possible. Label 3 was used in areas of high fissure uncertainty.

The manual segmentation provided an approximate estimate of fissure incompleteness with an average of $12.4\% \pm 8.3\%$ across the quantitative set. This was computed by considering the percentage of voxels labelled 2 and 3. Intra-observer variability was obtained by a repeated segmentation of 3 datasets with varying degrees of fissure incompleteness (6.7%, 23.0% and 31.3%). These were performed 14 days after to minimise recall bias. The intra-observer agreement for all lobar boundaries across all patients was $1.54\text{mm} \pm 0.45\text{mm}$.

²copdgene.org/sites/default/files/COPDGene%20MOP%2006.19.2009.pdf

IV. EXPERIMENTS & RESULTS

A. Fissure segmentation evaluation

We investigated the effects of parameters A and B (Eq. 2) and the performance of our segmentation framework on the quantitative set of dataset 2 using label 1 of the reference set.

We used the method presented by Xiao et al. [34] to evaluate our fissure segmentation. We did not define a volume of interest (VoI) using a 40mm width band around each reference as this ignores potential false positives in the validation. The F_1 -score was used as quantitative index of performance. It is defined as $2 \cdot (Precision \cdot Recall) / (Precision + Recall)$. The magnitude of F_1 reflects the similarity between the segmentation and the reference.

Precision and *Recall* are defined respectively as $TP_1 / (TP_1 + FP)$ and $TP_2 / (TP_2 + FN)$ [34].

Precision was quantified by considering the overlap of the binary result (\mathcal{S}) with the reference. A 3mm tolerance band was defined around the reference segmentation as performed by Xiao et al. [34]. Voxels of \mathcal{S} are classified as true positive (TP_1) if they fall within the 3mm band and false positive (FP) if otherwise. *Recall* was computed by defining a 3mm band around the binary result (\mathcal{S}). Reference voxels within this band were classified as TP_2 and those outside as false-negative (FN).

We segmented the fissures using parameters $A, B \in [0.05, 0.50]$. We illustrate the performance of the segmentation for a subset of the parameter $B \in [0.05, 0.10, 0.15, 0.25]$ with $A \in [0.05, 0.50]$ in Fig. 7. The best performance over all datasets was achieved with parameters $A = 0.25$ and $B = 0.10$ with a median F_1 -score of 0.90 with median False-Discovery Rate and False-Negative-Rate of 0.08 and 0.13 respectively. The F_1 -score remained relatively stable when set in the range $A^* = B^* \in [0.10, 0.35]$ (Fig. 7). The mean F_1 -score over all combinations ($A^* \times B^*$) was 0.87 ± 0.02 demonstrating stability in algorithm performance. The mean F_1 over all values of A for increasing values of B is 0.87 ± 0.03 (Fig 7a), 0.87 ± 0.03 (Fig 7b), 0.86 ± 0.04 (Fig 7c), 0.83 ± 0.08 (Fig. 7d). At higher values of B , this drops to 0.82 ± 0.09 ($B = 0.30$), 0.81 ± 0.10 ($B = 0.35$), 0.79 ± 0.12 ($B = 0.40$), 0.78 ± 0.14 ($B = 0.45$) and 0.77 ± 0.15 ($B = 0.50$). This is expected as higher values decreases the separation between the tissue and fissure-class distributions.

B. Lobe segmentation validation

Algorithm parameters quoted within Section II were used in the validation of dataset 1 and 2. The fissure filter parameters used were $A = 0.20$ and $B = 0.20$.

1) Dataset 1 - LOLA11—We evaluated our algorithm on the LOLA11 cohort and submitted our results for evaluation [32]. The LOLA11 evaluation metric is the volume overlap between the submission and the reference segmentation of one observer. The organisers defined a 2mm slack border around the borders of the lung and lobes to account for inter-observer variability. Voxels within this border were not accounted for during evaluation. The overlap is calculated for each lobe across all patients. We report the mean \pm standard deviation, first quartile (Q1), median and third quartile (Q3) of the scores across all 55 patients. The LOLA score is calculated as the average of all average overlaps over all

lobes. Table I shows the score for our lobe segmentation and those of van Rikxoort et al. [10] and Lassen et al. [7].

Five algorithms have been validated for lobe segmentation on this cohort. However, we have restricted our comparison to van Rikxoort et al. [10] and Lassen et al. [7] as they are fully automatic and do not require interactive post-processing to correct segmentations. We achieved the highest automatic average lobe score of 0.884 and an average median overlap of 0.950 (Table I).

2) Dataset 2 - Quantitative COPDGene—We assessed the performance of our algorithm quantitatively on dataset 2 using the mean, maximum and root-mean square error (RMSE) distance from the manual reference to the automatic boundary. This was performed by calculating the three-dimensional Euclidean distance between the reference voxels and the closest point on the automatic segmentation.

We assessed the segmentation for each label (Table IV). The algorithm achieved a mean of $2.01\text{mm} \pm 6.24\text{mm}$ when the fissures were visible (label 1). In cases where fissure extrapolation was possible (label 2), a mean of $5.16\text{mm} \pm 6.12\text{mm}$ was achieved. The performance dropped to $7.31\text{mm} \pm 4.88\text{mm}$ in regions of highest uncertainty (label 3). In one case, the right lobe segmentation failed and in a second case, segmentation of the right and left lobes failed. The failure was due to major errors in the airway branching labelling, which could not be corrected using our methodology. With a 2mm slack border, the mean distances were $1.65 \pm 3.28\text{mm}$ (label 1), $3.31 \pm 5.93\text{mm}$ (label 2) and $6.18 \pm 4.70\text{mm}$ (label 3). Errors were due to instances where the groupwise fissure was significantly different from label 2 and 3 of the reference or slightly biased the segmentation of label 1. Further typical errors were in cases where emphysametous bullae appearing as fissures lead to isolated errors in the lobe boundary segmentation.

The performance of our algorithm was also assessed against approximate fissure incompleteness (Fig. 8). The relationships between the mean distance and standard deviation of the distance to the closest points on the automatic segmentation were examined and the Pearson correlation coefficient was calculated for each lobar boundary. Weak relationships were observed between the mean distance and the degree of fissure incompleteness for each boundary (right minor: $\rho = .47$ ($p < 0.05$), right major: $\rho = .66$ ($p < 0.05$) and left major: $\rho = .35$ ($p > 0.05$)). Similar findings were observed in the standard deviation (right minor: $\rho = .46$ ($p < 0.05$), right major: $\rho = .61$ ($p < 0.05$) and left major: $\rho = .23$ ($p > 0.05$)).

3) Dataset 2 - Qualitative COPDGene—We qualitatively assessed (Table III) our algorithm on 50 patients with advanced disease using an adapted scoring system of van Rikxoort et al. [21], which scores the segmentations out of five. The radiologist assessed each lobe segmentation on the sagittal plane. The highest score (5) corresponded to a segmentation error below 3mm. A score of 4 reflected a segmentation error at any location between 3mm and 12mm. A score of 3 reflected a segmentation error greater than 12mm but where the overall lobe segmentation remained acceptable for analysis. The lowest scores (2 and 1) were awarded when the maximum segmentation error was greater than 12mm and

segmentation quality was either equivocal or unusable. Scores were assigned to both complete and incomplete fissures.

The algorithm showed good performance across complete (3.9 ± 0.3) and incomplete fissures (3.8 ± 0.5). There was one notable failure (score = 2) to segment the right major fissure in a case with a complete fissure and one failure (score=1) of a segmentation of a right major fissure when it was grossly incomplete. Across cases with complete and incomplete fissures, the most commonly awarded score was 4. Only a small proportion of cases had maximum errors $< 3\text{mm}$ (22% and 12.5% for complete and incomplete right minor fissures whilst 0% and 7.3% for complete and incomplete right major fissures).

C. Effect of the groupwise prior

We assessed the performance of the algorithm with and without the groupwise prior (Π) on dataset 2. We computed the mean of the distances from the automated segmentation to the reference and compared this to results using the prior. We did not include the boundaries that failed in the quantitative analysis of Section IV-B2 as these boundaries also failed without using the groupwise prior. We omitted cases with minor fissure incompleteness in the analysis. This was defined when a lobar boundary had less than 1% fissure incompleteness. This led to 19 analysed patients for the right major fissure, 23 for the right minor and 19 for the left major. There was $9.72\% \pm 8.66\%$ fissure incompleteness in the right major, $36.44\% \pm 18.91\%$ for the right minor and $11.61\% \pm 9.81\%$ for the left major in the new cohort. We calculated the cohort average for each boundary and for the segmentation labels of fissures not visible in CT, delineated as label 2 and 3. We performed a two- sample t-test under the null hypothesis that the mean results of the segmentation with and without the groupwise prior are significantly different.

We found a general increase in the distance to the reference for label 2 ($5.87\text{mm} \pm 3.72\text{mm}$ to $7.60\text{mm} \pm 6.49\text{mm}$) and label 3 ($7.10\text{mm} \pm 3.67\text{mm}$ to $8.59\text{mm} \pm 5.77\text{mm}$) when excluding the groupwise prior. We did not find a significant difference between the sets of mean distances for each lobar boundary stratified by reference label (Table II). This is due to the fact there may be extreme differences due to failures without the prior, smaller improvements using the prior but also cases where the prior negatively affects extrapolation of the fissure.

In areas of significant fissure incompleteness, the groupwise prior may help avoid leaking of the seed labels during the surface fitting whilst guiding the segmentation to the most probable location based on the population and the patient anatomy. This occurred in several cases (Figure 11(a-f)) where either the left major border or the right major border failed without the prior. Within this cohort, the right minor fissure had the highest level of fissure incompleteness. In various cases (Figure 11 (g-i)), the prior helped drive the lobar border towards the reference. However, there are several modes of variation in the right minor fissure (Figure 5). The patient anatomy may differ greatly from the population mean. The prior may negatively affect the final segmentation in areas of incomplete fissures. This led to a smaller difference in the population means in the right minor fissure (6.28mm to 6.51mm in label 2 and 6.41mm to 6.50mm in label 3). Despite this limitation, we can conclude that

the groupwise fissure prior, constructed from the same cohort using a combination of complete and incomplete segmented fissures is advantageous.

V. DISCUSSION AND CONCLUSION

We have presented a novel lobe segmentation algorithm based on an unsupervised segmentation of the fissures with iterative false-positive removal, the creation of a groupwise fissure prior and a cost function combining patient and population information. Our algorithm does not require prior training or manual labelling to segment the fissures and build a population prior of the fissures. Fissure probabilities were obtained by parameterising a fissure enhancement filter with a Gaussian Mixture Model (GMM). Smoothness and curvature constraints were enforced in the segmentation by considering a Markov Random Field (MRF) regularisation. This led to rejection of most false-positives leading to high maximum F_1 -score of 0.90. A method to construct a groupwise fissure prior given complete and incomplete fissures in a population was presented. We evaluated its role in identifying incomplete fissures whilst minimising potential segmentation failures. The method was validated on 55 cases from the LOLA11 study [32] and on 80 datasets from the COPDGene study [24]. We illustrated its applicability in correctly segmenting the lobes of patients with varying levels of disease severity and fissure incompleteness.

Correct removal of false-positive fissures is necessary to accurately segment the fissures. Whilst supervised filter techniques have been utilised to segment the fissures [8], [9], [12], they require a training dataset to perform the classification. Manual labelling of voxels is laborious and may not always be practicable in a clinical setting. The fissure segmentation used by van Rikxoort et al. [22] required a training set as part of their algorithm and can only be employed when manual labelling is possible by an expert. Moreover, the applicability of a training set built on an independent set of scans applied to those acquired on different scanners is debatable.

Ross et al. [12] exploited a deformable model to identify the fissure surfaces, which may fail when the patient anatomy cannot be modelled by the atlas. The fissure enhancement applied by Lassen et al. [7] required experimentation to yield optimal ranges for the Hessian eigenvalues. Applying pre-existing thresholds to new datasets and those acquired at lower doses can be problematic and may cause undesirable drops in algorithm specificity. Applying hard-constraints on eigenvalue magnitudes in new scans may not be beneficial. The eigenvalue range may differ whilst the ratio is expected to remain constant. This is because the ratio will model different orientation patterns unique to various structures (spherical, tubular and sheet-like). This may cause an over-segmentation with too many false positives. The limitation of pre-existing threshold can also be applied to the filter of Wiemker et al. [19], who developed a weighting term based on the expected intensity of the fissures. It requires specific knowledge about the HU of the fissures to determine parameters unique to scans. Finding optimal parameters that yield a robust filter across a broad range of datasets is difficult and not desirable.

Our segmentation framework requires little prior knowledge regarding algorithm parameters and will be more robust than methods requiring prior training. Our fissure enhancement

filter is based on ratios of Hessian eigenvalues. Given a set of new scans, separability between class distributions will exist since relationships between the eigenvalues should remain constant. Since we learn model parameters of the GMM, the segmentation of the fissures will be flexible when processing new scans. Classifying voxels by learning the underlying class distributions will capture a range of filter values; which can vary for each dataset. The integration of the MRF regularisation increases the robustness of the segmentation to noise. A voxel with a poor filter response can still receive a high fissure class probability when considering neighbourhood constraints on pairwise probabilities and Hessian eigenvectors.

We showed in Section IV-A that our method achieved a high median F_1 -score of 0.90 and insensitivity to the filter input parameters. These parameters govern the separability of the class distributions and do not rely on knowledge of specific CT features. There is a minor dependence on the initialisation of the model parameters in the GMM. However, the iterative framework for increasing the MRF regularisation means no user-interaction is required. The iterative increase also deals with false-positive rejection as constraints on neighbourhood properties are given more weight until algorithm convergence. Performance of the fissure segmentation could be improved in future work by modelling the signal as a mixture of skew-normals akin to the work of Hämäläinen et al. [35] since we chose a GMM for mathematical simplicity.

We evaluated our algorithm on the LOLA11 dataset [32], which enabled direct comparison with the work of Lassen et al. [7] and van Rikxoort et al. [10]. We achieved the highest score of 0.884 in comparison to Lassen et al. [7] (0.881) and van Rikxoort et al. [10] (0.851). Both algorithms used superior lung segmentation algorithms (0.947 (our method) versus 0.962 [10] and 0.971 [7]), which may have had detrimental effects on our lobe segmentation scores in the most challenging cases (e.g. Fig. 10h, i, k and l).

The quantitative experiment on dataset 2 highlighted the accuracy of the algorithm in areas with varying fissure visibility (Table IV). The high standard deviation associated with the segmentation of the right minor fissure was due to the failure of lobar segmentation in 2/30 cases. In these two cases, post-processing of airway labelling errors could not be automatically corrected. The respective fissure means and standard deviations were: 1.52mm \pm 1.49mm, 46.06mm \pm 30.74mm and 1.33mm \pm 2.06mm for the right major and minor fissures and left major fissure respectively, which displays the isolated error. The large maximum errors in certain cases with low mean distances occurred in isolated areas close to the ribcage and near the lung hilum where the automated segmentation disagreed significantly with the radiologist.

Qualitative testing (Table III) highlighted the ability of the algorithm to produce good segmentation results in cases with higher severities of disease. The low proportion of lobar boundaries scoring 5 (errors < 3 mm) in cases with complete fissures (Table IIIa) are a result of the narrow boundary definitions of the scoring system. The scoring system may not adequately reflect the performance of the algorithm since an isolated error will reduce the score to 4 when it would otherwise be graded as 5. Most of these errors were less than 6mm from the reference fissure line and occurred in isolated regions prone to artefacts such as

close to the rib-cage, the lung hilum and the intersection between the right major and minor boundaries. Importantly, the algorithm was able to interpolate incomplete fissures (Table IIIb) and demonstrated equivalence in performance to cases with complete fissures. The findings highlight the ability of the groupwise prior in conjunction with information from the vessel and airway tree to successfully guide fissure segmentation towards correct locations as defined by the reference standard.

Our work bears many similarities with the implementation of Lassen et al. [7]. They also exploited information from auxiliary structures by combining the airway and vessel tree with the segmented fissures to create a cost image for watershed segmentation. Our algorithm differs primarily in the fissure segmentation and in the inclusion of population information in the cost image. Priors dependent on the segmentation of the vessel and airway tree might not always be fully informative. Airway tree segmentation is challenging and may not be sufficiently segmented to provide enough information about the location of lobar borders. The vessel tree may also not provide sufficient information in areas of largely incomplete fissures. The additional information provided by the groupwise fissure prior helps mitigate these issues. Within the LOLA11 dataset, the effect is marginal on the overall scores (0.884 versus 0.881). The dataset included many highly irregular scans, which made it difficult to create an accurate population model. The effect of the groupwise prior was more noticeable in dataset 2 from the COPDGene study (Fig. 6 and 11) where we demonstrated the utility of the groupwise fissure.

The technique by which we construct the groupwise prior has the advantage of not requiring any pre-existing data. This does not require an expert to manually delineate complete fissures including visible and non-visible fissures. Since we construct the prior on the current set of data, it is not biased towards particular types of imaging data. The groupwise prior acted as a guide or region of confidence within the patient space rather than rigidly guiding the segmentation based on shapes in the training set, facilitating the segmentation of lobes of varying shape. Its effect was demonstrated on dataset 2 (Table II) and visualised in Fig. 11. Our method for constructing the groupwise prior suffers from a lack of flexibility in comparison to deformable models. A prior created by averaging all segmented fissures in the groupwise space may be over simplistic. Despite the simplicity of its construction, the application can bias the results in certain cases. This was seen in the weak relationships calculated between segmentation errors and fissure incompleteness (Fig. 8) but also when quantifying the effect of the prior (Section IV-C). Since the average fissure is directly added within the patient space, it does not take into account the shape of the segmented fissure. This introduces a bias if the patient anatomy differs significantly from the mean. This occurred mostly in the right minor fissure, where several modes in the population exist (Fig. 5). This led to a smaller average increase in the errors when testing algorithm performance without the prior (Table II).

The dependence of our work on the construction of a groupwise space is a limitation. Groupwise registration is computationally expensive and measuring registration accuracy of inter-patient registration is difficult. Errors in the registration may be present, which can decrease the strength of our calculated groupwise prior. The work of Li et al. [36] used annotated landmarks from the airway tree to drive inter-patient landmark and intensity-based

registration, indicating the applicability of detecting landmarks in inter-patient registration. Inclusion of the vessel density map in a multi-modal registration scheme to construct our groupwise space may then produce a more accurate fissure prior. Improving the inter-patient registration is then likely to increase the flexibility of the fissure prior. The transformation between the average space and the patient space acts as a deformable model. A more accurate mapping; obtained by including extra morphological information will help deform the groupwise prior to more unusual geometries.

The need to segment the airway tree may also decrease the applicability of our framework. Segmentation of the bronchial tree is an important determinant in the success of our algorithm as demonstrated by several failures in dataset 2. The vessel density map produces a good approximation of the fissure location, which may negate the need to use the airway tree in the fissure segmentation and cost function. However, initial seed labelling for the watershed still relies on the airway tree. Whilst we developed post-processing methods to make the method more robust to the quality of the airway segmentation, errors can lead to failures. A combination of the groupwise framework and the labelling method described in Section II-D3 may help generate better seeds without needing the airway tree.

The segmentation of incomplete fissures remains one of the biggest challenges in lobe segmentation. We presented the simultaneous construction of a groupwise prior to address this challenge. When the fissures are not-visible on CT, this is because they may be congenitally absent or destroyed by inflammatory disease processes. The segmentation may therefore be creating an artificial division between lobes. In reality, the anatomical boundary between the lobes has either been destroyed or is absent. This is seldom mentioned in the lobe segmentation literature. When comparing our results with the label 3 reference, it is therefore important to note we are comparing algorithm extrapolation with the educated guess of an expert. Furthermore, it is not yet known what accuracy is needed in the segmentation of incomplete fissures to produce regional markers of disease that are clinically useful.

Despite the ability to correctly guide the segmentation in regions of incomplete fissures in most cases, the application of the groupwise prior requires further work. There may be issues when the mean of the population deviates significantly from the patient being segmented. In order to fully exploit the power obtained by fusing complete and incomplete segmentations, it is necessary to dynamically weight the groupwise prior in regions when it is needed and regions where information stemming from the patient is sufficient. Another solution may lie in creating various fissure models using different sets of patients from the population to mimic multi-atlas selection.

In conclusion, we have presented a lobe segmentation algorithm, which requires no prior training or manual labelling to both segment the fissures and build a population prior of the fissures. We have tested the method on 135 different datasets with varying levels of disease severity and complexity. The presented algorithm can be used in large studies to perform accurate regional quantification of disease progression and shows great promise to be integrated within a clinical setting.

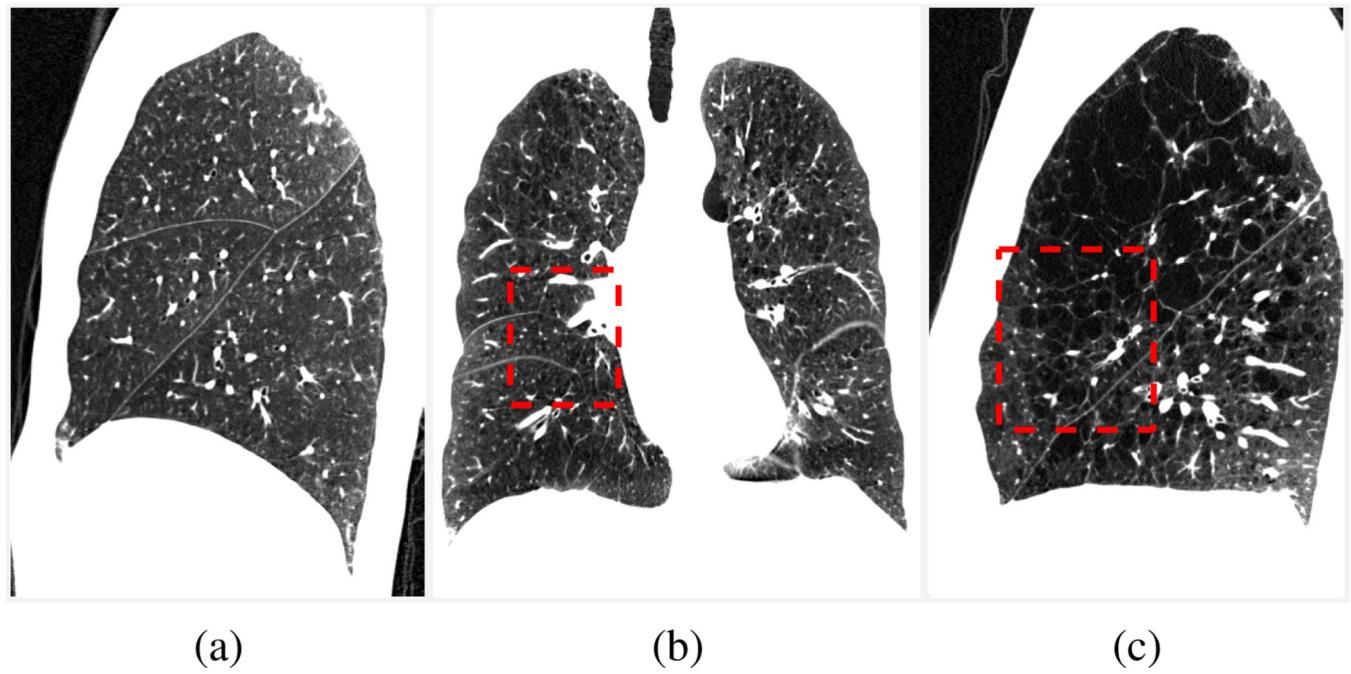
Acknowledgments

This work was supported by the EPSRC (EP/H046410/1 and EP/K502959/1). Additional support was provided by the UCLH NIHR RCF Senior Investigator Award (RCF107/DH/2014). This research used data (phs000179.v5.p2) generated by the COPDGene study, which was supported by NIH grants U01HL089856 and U01HL089897. The authors would like to thank the help of Dr. Tom Doel, Jorge Cardoso and Marc Modat for their helpful guidance.

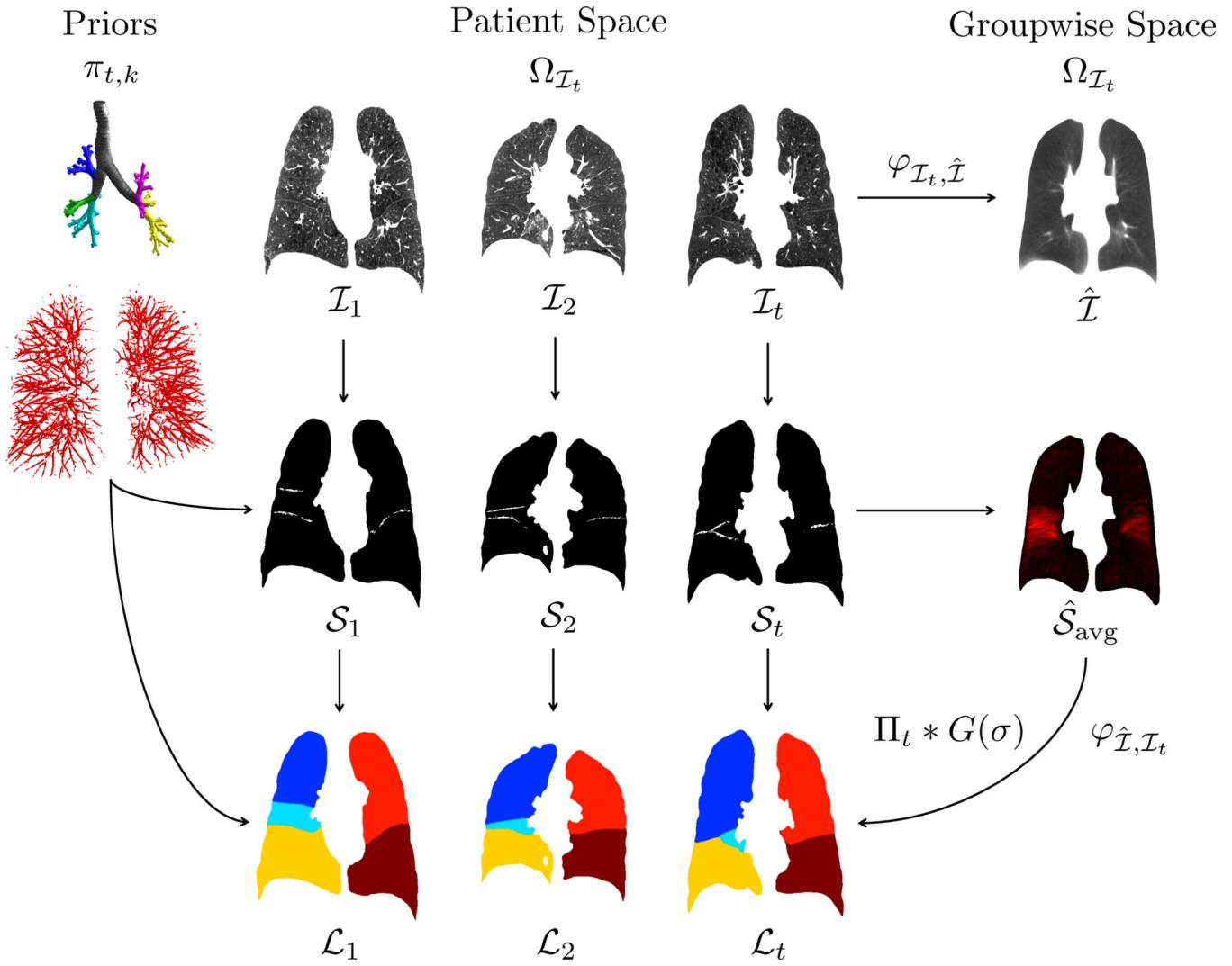
References

1. Tanabe N, Muro S, Tanaka S, Sato S, Oguma T, Kiyokawa H, Takahashi T, Kinose D, Hoshino Y, Kubo T, Ogawa E, Hirai T, Mishima M. Emphysema distribution and annual changes in pulmonary function in male patients with chronic obstructive pulmonary disease. *Respiratory research*. 2012; 13:31. [PubMed: 22512922]
2. Weder W, Thurnheer R, Stammberger U, Bürge M, Russi EW, Bloch KE. Radiologic emphysema morphology is associated with outcome after surgical lung volume reduction. *Annals of Thoracic Surgery*. 1997; 64(2):313–320. [PubMed: 9262567]
3. Cederlund K, Tylén U, Jorfeldt L, Aspelin P. Classification of emphysema in candidates for lung volume reduction surgery: A new objective and surgically oriented model for describing CT severity and heterogeneity. *Chest*. 2002; 122(2):590–596. [PubMed: 12171837]
4. Hayashi H, Nagaoki K, Otsuji H. Radiographic and CT Major Fissures 1. *Radiographics*. 2001;861–874. [PubMed: 11452059]
5. Aziz A, Ashizawa K, Nagaoki K, Hayashi K. High resolution CT anatomy of the pulmonary fissures. *Journal of thoracic imaging*. 2004; 19(3):186–191. [PubMed: 15273615]
6. Doel T, Gavaghan DJ, Grau V. Review of automatic pulmonary lobe segmentation methods from CT. *Computerized Medical Imaging and Graphics*. 2015; 40:13–29. [PubMed: 25467805]
7. Lassen E, van Rikxoort B. Automatic segmentation of the pulmonary lobes from chest CT scans based on fissures, vessels, and bronchi. *IEEE transactions on medical imaging*. 2013; 32(2):210–22. [PubMed: 23014712]
8. Lo P, van Rikxoort EM, Abtin F, Ahmad S, Ordoonkhani A, Goldin J, Brown MS. Automated segmentation of pulmonary lobes in chest CT scans using evolving surfaces. 2013; 8669
9. van Rikxoort EM, van Ginneken B, Klik M, Prokop M. Supervised enhancement filters: application to fissure detection in chest CT scans. *IEEE transactions on medical imaging*. 2008; 27(1):1–10. [PubMed: 18270056]
10. van Rikxoort EM, Prokop M, de Hoop B, Viergever MA, Pluim JPW, van Ginneken B. Automatic segmentation of pulmonary lobes robust against incomplete fissures. *IEEE transactions on medical imaging*. 2010; 29(6):1286–96. [PubMed: 20304724]
11. Schmidt-Richberg A, Ehrhardt J, Werner R, Handels H. Lung registration with improved fissure alignment by integration of pulmonary lobe segmentation. *Medical image computing and computer-assisted intervention (MICCAI)*. 2012; 15(Pt 2):74–81.
12. Ross JC, Kindlmann GL, Okajima Y, Hatabu H, Díaz Aa, Silverman EK, Washko GR, Dy J, San José Estépar R. Pulmonary lobe segmentation based on ridge surface sampling and shape model fitting. *Medical physics*. 2013; 40(12):121903. [PubMed: 24320514]
13. Zhang L, Hoffman EA, Reinhardt JM. Atlas-driven lung lobe segmentation in volumetric X-ray CT images. *IEEE transactions on medical imaging*. 2006; 25(1):1–16. [PubMed: 16398410]
14. Doel T, Matin TN, Gleeson FV, Gavaghan DJ, Grau V. Pulmonary lobe segmentation from CT images using fissureness, airways, vessels and multilevel B-splines. 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI). 2012:1491–1494.
15. Ukil JM, Reinhardt S. Anatomy-guided lung lobe segmentation and fissure analysis in x-ray CT images. *IEEE transactions on medical imaging*. 2009; 28(2):202–214. [PubMed: 19188109]
16. Zhou X, Hayashi T, Hara T, Fujita H, Yokoyama R, Kiryu T, Hoshi H. Automatic segmentation and recognition of anatomical lung structures from high-resolution chest CT images. *Computerized Medical Imaging and Graphics*. 2006; 30(5):299–313. [PubMed: 16920331]

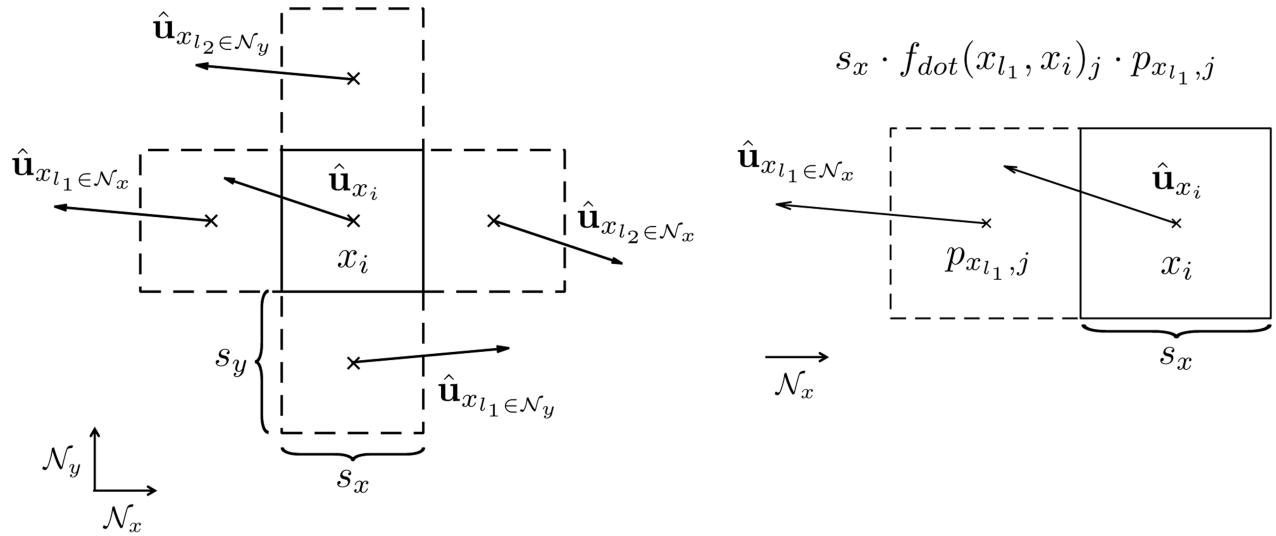
17. Pu J, Zheng B, Leader JK, Fuhrman C, Knollmann F, Klym A, Gur D. Pulmonary lobe segmentation in CT examinations using implicit surface fitting. *IEEE Transactions on Medical Imaging*. 2009; 28(12):1986–1996. [PubMed: 19628453]
18. Nimura Y, Kitasaka T, Honma H, Takabatake H, Mori M, Natori H, Mori K. Lung lobe segmentation based on statistical atlas and graph cuts. *SPIE Proceedings*. 2012; 8315
19. Wiemker R, Blow T, Blaffert T. Unsupervised extraction of the pulmonary interlobar fissures from high resolution thoracic CT data. *International Congress Series*. 2005; 1281:1121–1126.
20. Klinder T, Wendland H, Wiemker R. Lobar fissure detection using line enhancing filters. *SPIE Medical imaging*. 2013; 8669:86693C.
21. van Rikxoort B, van Ginneken EM. Automated segmentation of pulmonary structures in thoracic computed tomography scans: a review. *Physics in medicine and biology*. 2013; 58(17):R187–220. [PubMed: 23956328]
22. Van Rikxoort EM, Prokop M, De Hoop B, Viergever Ma, Pluim JPW, Van Ginneken B. Automatic segmentation of pulmonary lobes robust against incomplete fissures. *IEEE Transactions on Medical Imaging*. 2010; 29(6):1286–1296. [PubMed: 20304724]
23. Zhang L, Hoffman EA, Reinhardt JM. Atlas-driven lung lobe segmentation in volumetric X-ray CT images. *IEEE transactions on medical imaging*. 2006; 25(1):1–16. [PubMed: 16398410]
24. Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, Curran-Everett D, Silverman EK, Crapo JD. Genetic epidemiology of COPD (COPDGene) study design. *COPD*. 2010; 7(1):32–43. [PubMed: 20214461]
25. Hu S, Hoffman EA, Reinhardt JM. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. *IEEE transactions on medical imaging*. 2001; 20(6):490–8. [PubMed: 11437109]
26. Frangi, MaVRo F, Niessen, Wiro J., Vincken, Koen L. Multiscale vessel enhancement filtering. *Medical image computing and computer-assisted intervention (MICCAI)*. 1998; 1496:130–137.
27. Van Leemput K, Maes F, Vandermeulen D, Suetens P. Automated model-based tissue classification of MR images of the brain. *IEEE transactions on medical imaging*. 1999; 18(10):897–908. [PubMed: 10628949]
28. Cardoso MJ, Clarkson MJ, Ridgway GR, Modat M, Fox NC, Ourselin S. LoAd: A locally adaptive cortical segmentation algorithm. *NeuroImage*. 2011; 56(3):1386–1397. [PubMed: 21316470]
29. Modat M, Ridgway GR, Taylor ZA, Lehmann M, Barnes J, Hawkes DJ, Fox NC, Ourselin S. Fast free-form deformation using graphics processing units. *Computer methods and programs in biomedicine*. 2010; 98(3):278–84. [PubMed: 19818524]
30. Ashburner, J. Ph.D. dissertation. 2000. Computational neuroanatomy.
31. Modat M, Cash DM, Daga P, Winston GP, Duncan JS, Ourselin S. Global image registration using a symmetric block-matching approach. *Journal of Medical Imaging*. Sep. 2014 1(2):024003. [PubMed: 26158035]
32. LOLA11. [Online]. Available: <https://grand-challenge.org/site/lola11/>
33. Yushkevich PA, Piven J, Cody HH, Smith RG, Ho S, Gee JC, Gerig G. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*. 2006; 31(3):1116–28. [PubMed: 16545965]
34. Xiao C, Stoel B, Bakker M, Peng Y, Stolk J, Staring M. Pulmonary Fissure Detection in CT Images Using a Derivative of Stick Filter. *IEEE Transactions on Medical Imaging*. 2016; 35(6):1488–1500. [PubMed: 26766371]
35. Hame Y, Angelini ED, Hoffman EA, Barr RG, Laine AF. Adaptive quantification and longitudinal analysis of pulmonary emphysema with a hidden Markov measure field model. *IEEE transactions on medical imaging*. 2014; 33(7):1527–40. [PubMed: 24759984]
36. Li B, Christensen GE, Hoffman EA, McLennan G, Reinhardt JM. Establishing a normative atlas of the human lung: intersubject warping and registration of volumetric CT images. *Academic radiology*. Mar; 2003 10(3):255–65. [PubMed: 12643552]

**Fig. 1.**

Example in the variation of fissure integrity. Scans with complete (a) and incomplete (b and c) fissures can be visualised. The incomplete fissures are due to fusion of lung tissue (b) or pathology (c). Slices are displayed in the intensity range $\mathcal{Q} \in [-1024 \text{ HU}, -600 \text{ HU}]$.

**Fig. 2.**

Lobe segmentation algorithm for processing a set of T patients, given by $\mathcal{Q} = \{\mathcal{Q}_1, \mathcal{Q}_2, \dots, \mathcal{Q}_t, \dots, \mathcal{Q}_T\}$. The lung mask, airway and vascular tree are segmented as preprocessing steps. Anatomical information ($\pi_{t,k}$, $k = (\text{fissure, tissue})$) is derived from the airway and vascular tree. A probabilistic segmentation of the fissures based on a filter (\mathcal{F}_t) exploits these priors. This yields fissure segmentations (\mathcal{S}_t) for each patient t . For a given cohort ($\mathcal{Q} = \{\mathcal{Q}_1, \mathcal{Q}_2, \dots, \mathcal{Q}_T\}$), a groupwise space is constructed yielding the set of transformations $\varphi_{\mathcal{Q}_t, \hat{\mathcal{Q}}}$ to the common space $\hat{\mathcal{Q}}$. This space is exploited to construct an average model of the fissures ($\hat{\mathcal{S}}_{\text{avg}}$). This is combined with the segmented fissures \mathcal{S}_t and the patient-specific anatomical information ($\pi_{t,k}$) in a cost function for non-parametric surface fitting. Final lobe segmentations ($\mathcal{L} = \{\mathcal{L}_1, \mathcal{L}_2, \dots, \mathcal{L}_T\}$) are obtained for each patient (t) in the cohort ($\mathcal{Q} = \{\mathcal{Q}_1, \mathcal{Q}_2, \dots, \mathcal{Q}_T\}$).



(a) Two-dimensional neighbourhood \mathcal{N}_i (b) Features used in the MRF energy for a component (l_1) in \mathcal{N}_i^x

Fig. 3.

Illustration of the local neighbourhood \mathcal{N}_i used in Eq. 13. The index l iterates over each component in \mathcal{N}_i whilst j is the tissue class iterator. The real-world distances are represented by s_x and s_y , both measured in mm.

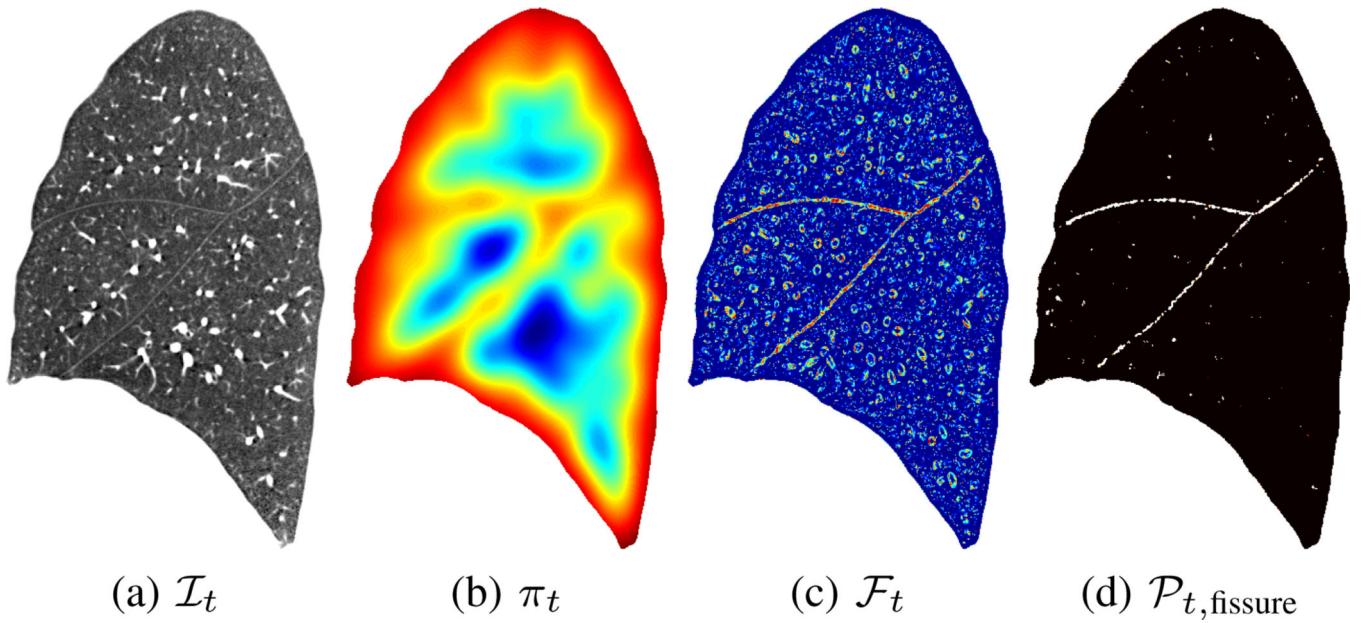


Fig. 4.

Given a patient \mathcal{Q}_t (a), an anatomical prior (π_t) (b) is derived from the vessel density and airway tree distance transform. A multi-scale filter is applied to \mathcal{Q}_t to yield \mathcal{F}_t (c). Gaussian mixture modelling with an MRF yields a probabilistic segmentation (d) of the fissures $\mathcal{P}_{t,\text{fissure}}$.

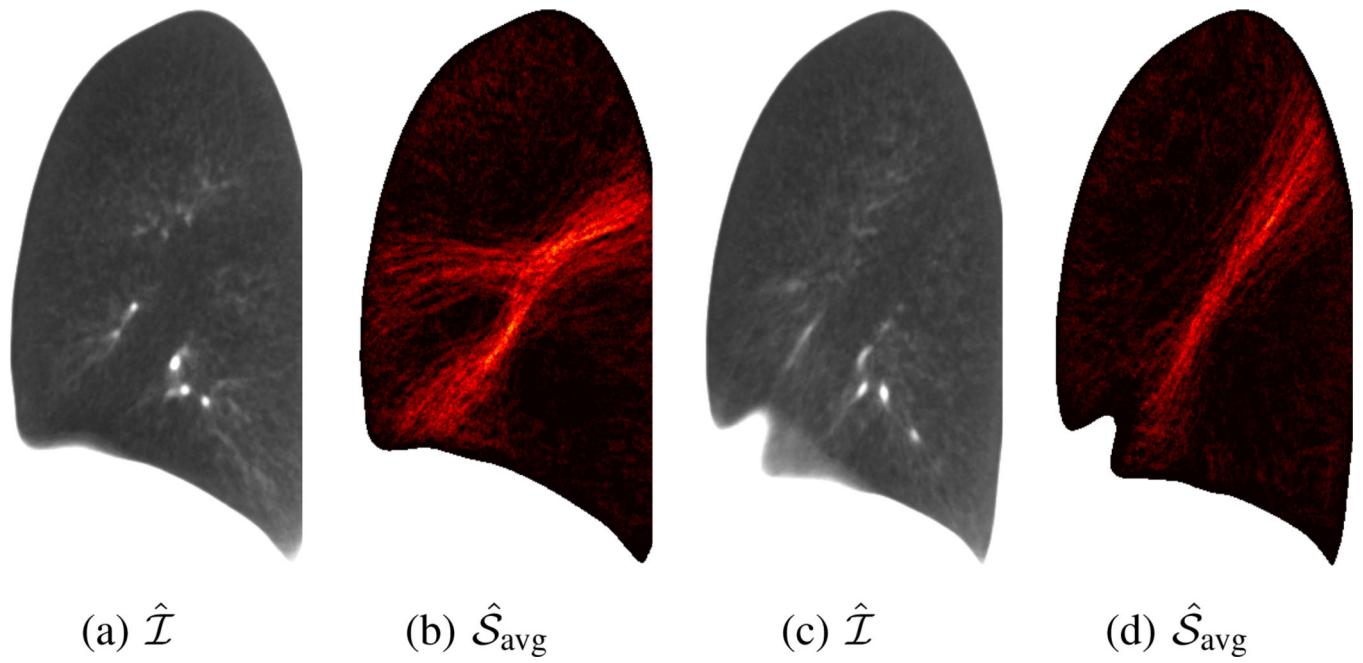


Fig. 5.

Average fissure in the groupwise space $\Omega_{\hat{\mathcal{Q}}}$. The average lung at the sagittal midsection of the right and left lung with the respective average fissures can be seen.

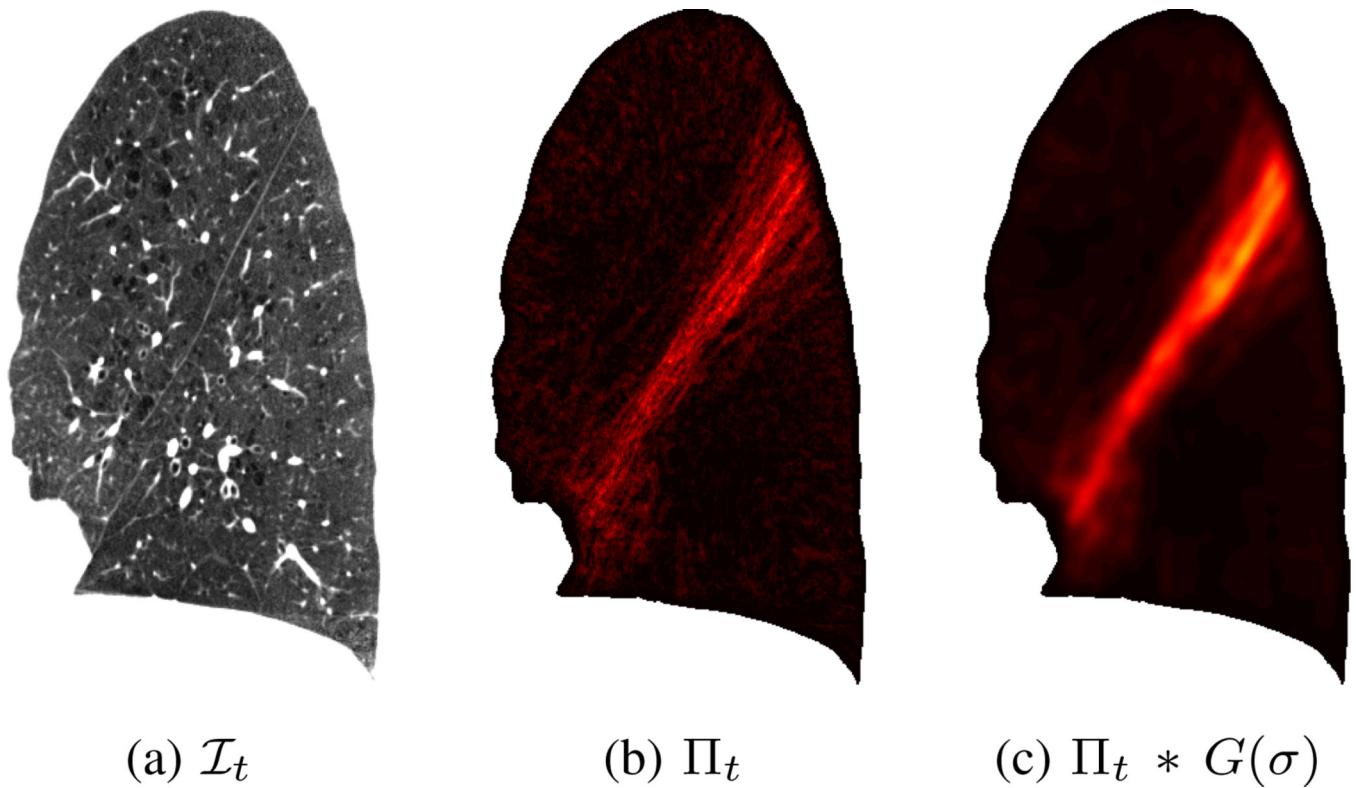


Fig. 6.

Groupwise prior in the patient space $\Omega_{\mathcal{Q}_t}$ a) The patient volume (\mathcal{Q}_t), b) the resampled average fissure (Π_t) and c) the smoothed prior ($\Pi_t * G(\sigma)$).

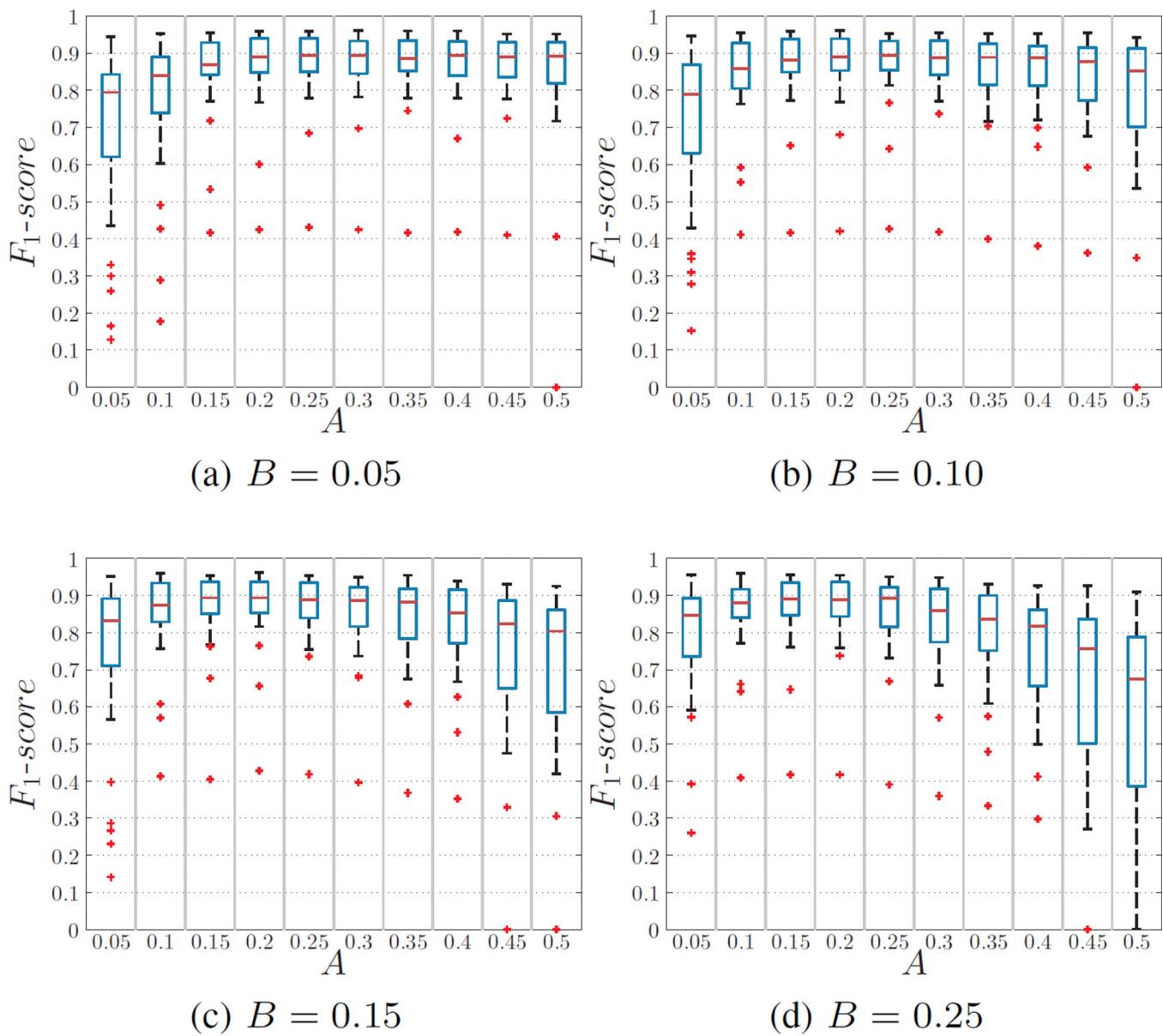
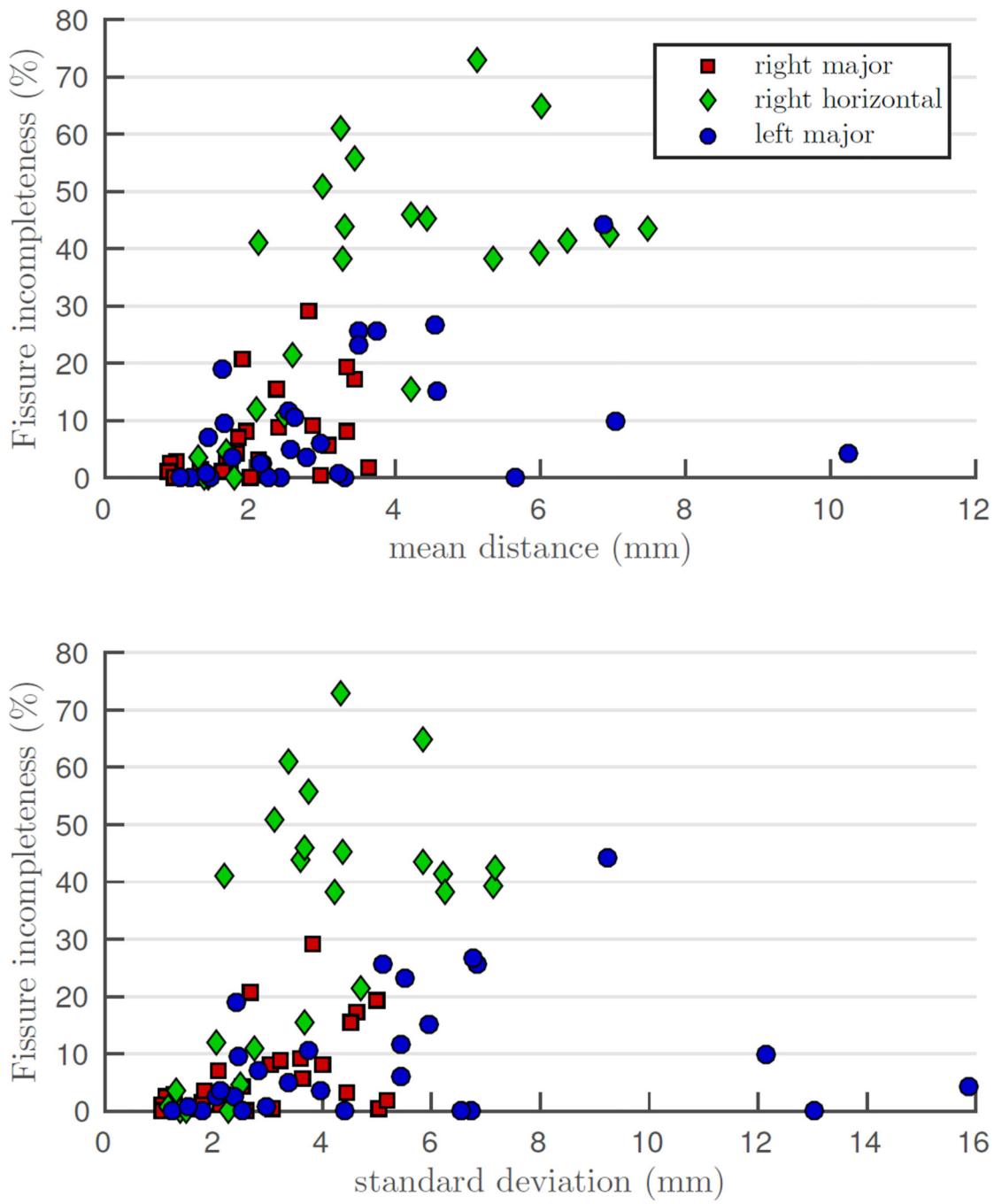


Fig. 7.
Boxplots of the fissure segmentation with increasing values of B whilst varying constant A of the multiscale filter (Eq. 2).

**Fig. 8.**

Relationship between incompleteness and segmentation performance. One case where the right major fissure failed and another where both the left and right lungs failed were removed as outliers for better visualisation.

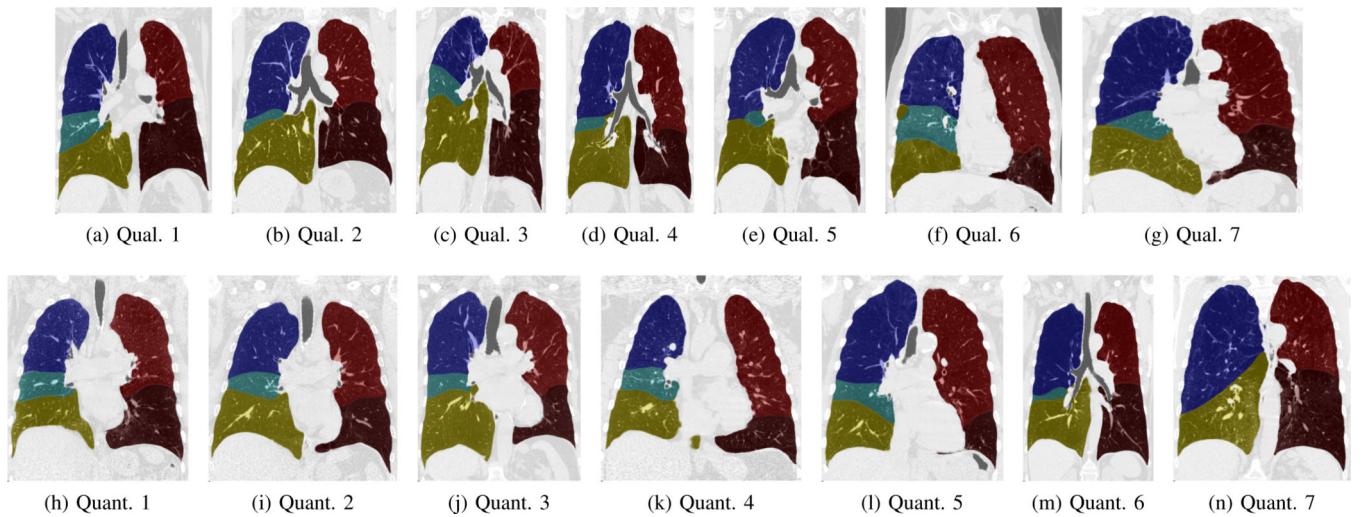


Fig. 9.

Illustration of the segmentation on examples from the qualitative set of dataset 2 (top row) and the quantitative set (bottom row) at various levels of COPD severity.

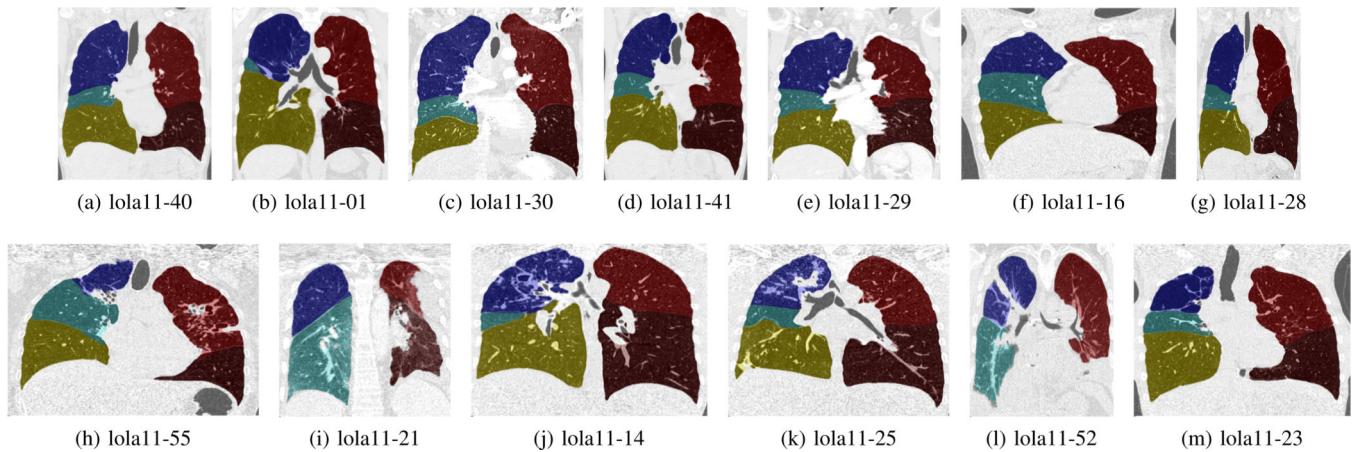
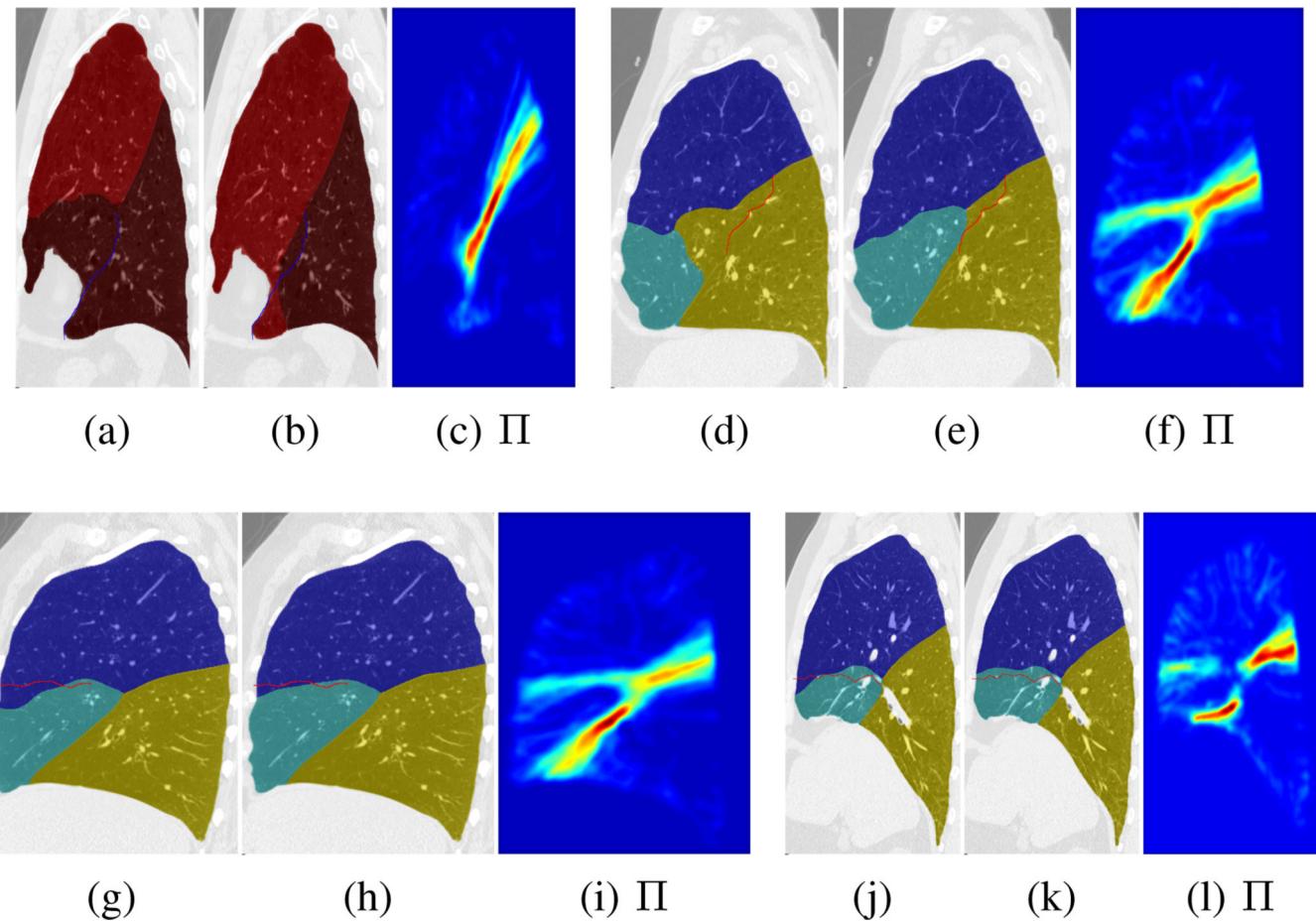


Fig. 10.

Illustration of the segmentation on a variety of cases from dataset 1 with complete and incomplete fissures in addition to various levels of pathology.

**Fig. 11.**

Lobe segmentation results without and with the groupwise prior (Π) in four different patients. The reference for non-visible fissures (label 2 and 3) is overlaid on the lobe segmentation. The colour of the reference was chosen to aid the visualisation and is not representative of the segmentation label.

Table I

LOLA 11 lobe segmentation overlap results

Lobe	Mean \pm STD	Q1	Median	Q3
LUL	0.906 \pm 0.202	0.946	0.975	0.988
LLL	0.880 \pm 0.243	0.919	0.962	0.980
RUL	0.928 \pm 0.071	0.888	0.960	0.980
RML	0.799 \pm 0.235	0.759	0.891	0.941
RLL	0.908 \pm 0.194	0.937	0.961	0.976
Our method	0.884		0.950	
Lassen et al. [7]	0.881		0.951	
van Rikxkoort et al. [10]	0.851		0.943	

Table II

Quantitative Results - Fissure Metrics with and without Groupwise Prior

		WITH (mm)	WITHOUT (mm)	p-Value
Label 2	Right Major	5.29 ± 3.00	5.71 ± 2.75	p=0.65
	Right Minor	6.28 ± 3.52	6.51 ± 3.51	p=0.83
	Left Major	6.05 ± 4.64	10.59 ± 16.32	p=0.25
	Total	5.87 ± 3.72	7.60 ± 6.49	-
Label 3	Right Major	6.94 ± 3.24	8.18 ± 4.28	p=0.34
	Right Minor	6.41 ± 2.81	6.50 ± 2.84	p=0.92
	Left Major	7.94 ± 4.95	11.09 ± 10.19	p=0.67
	Total	7.10 ± 3.67	8.59 ± 5.77	-

Table III

Qualitative Scoring Results for Complete Fissures (a) and Incomplete Fissures (b)

Score	Right Major	Right Minor	Left Major	Total Score
% 5	0.0	22.0	0.0	
% 4	82.4	58.0	76.5	
% 3	15.6	20.0	23.5	3.9 ± 0.3
% 2	2.0	0.0	0.0	
% 1	0.0	0.0	0.0	

(a)

Score	Right Major	Right Minor	Left Major	Total Score
% 5	7.3	12.5	2.4	
% 4	68.3	50.0	70.7	
% 3	22.0	37.5	26.8	3.8 ± 0.5
% 2	0.0	0.0	0.0	
% 1	2.4	0.0	0.0	

(b)

Table IV

Quantitative Results - Fissure Metrics

		average of patient mean (mm)	average of patient max (mm)	average of patient rmse (mm)
Label 1	Right Major	1.66 \pm 3.61	15.42 \pm 14.91	2.72 \pm 5.67
	Right Minor	2.31 \pm 11.10	18.60 \pm 20.10	3.32 \pm 12.82
	Left Major	2.07 \pm 4.03	28.50 \pm 27.89	3.95 \pm 6.85
	Total	2.01 \pm 6.24	20.85 \pm 20.96	3.33 \pm 8.44
Label 2	Right Major	5.44 \pm 5.55	14.22 \pm 14.45	6.69 \pm 6.98
	Right Minor	5.39 \pm 8.50	18.65 \pm 18.96	7.23 \pm 8.48
	Left Major	4.65 \pm 4.30	16.15 \pm 16.75	6.01 \pm 5.81
	Total	5.16 \pm 6.12	16.34 \pm 16.72	6.64 \pm 7.09
Label 3	Right Major	7.21 \pm 4.18	20.95 \pm 21.02	8.65 \pm 6.15
	Right Minor	7.56 \pm 5.49	19.05 \pm 17.62	8.44 \pm 7.28
	Left Major	7.16 \pm 34.96	23.23 \pm 23.23	9.78 \pm 5.58
	Total	7.31 \pm 4.88	21.08 \pm 20.62	8.95 \pm 6.34