Analysis of airways in computed tomography

Petersen, Jens

Publication date:
2014

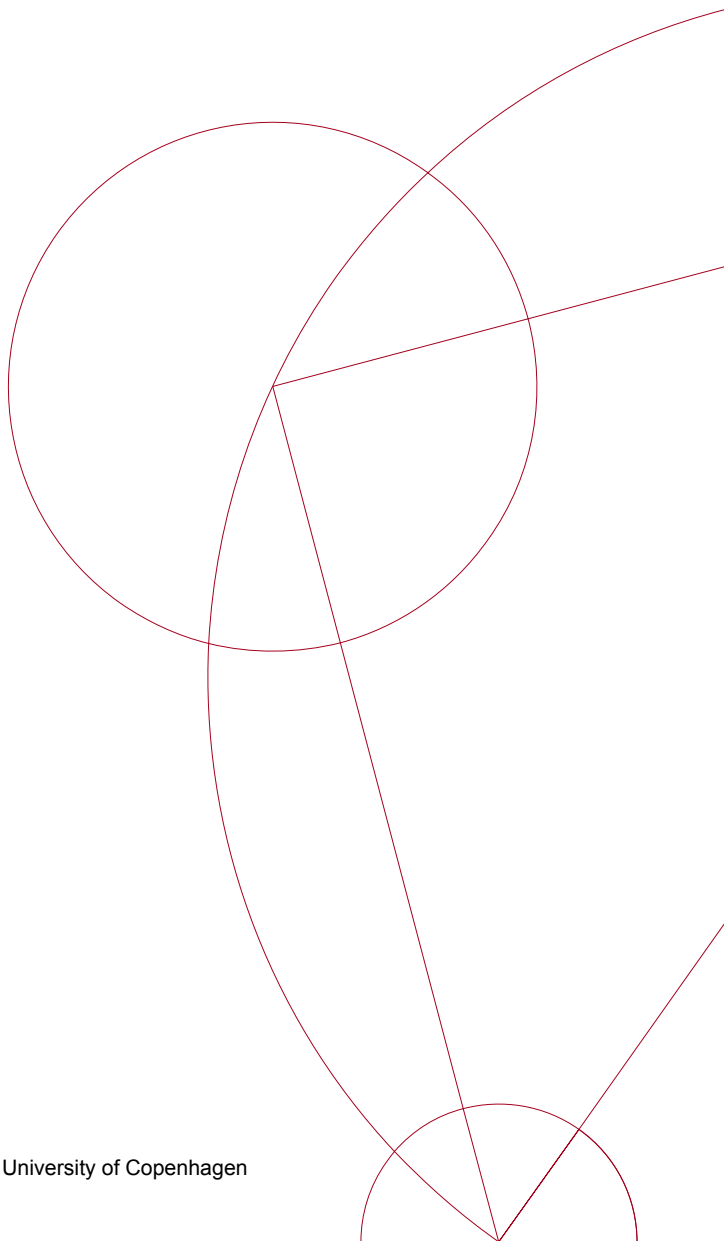
Document version
Publisher's PDF, also known as Version of record

FACULTY OF SCIENCE
UNIVERSITY OF COPENHAGEN

PhD thesis

Jens Petersen

Analysis of Airways in Computed Tomography



Academic advisor: Marleen de Bruijne

Co-supervisor: Asger Dirksen

Submitted: 25/02/2014

This thesis has been submitted to the PhD School of The Faculty of Science, University of Copenhagen

Abstract

Computed Tomography (CT) allows un-occluded three-dimensional views of the interior of scanned objects. In medical applications this offers the opportunity to diagnose or study diseases, such as Chronic Obstructive Pulmonary Disease (COPD) with limited consequence to the patients. COPD is major cause of death and disability world-wide. It affects how lungs function through two competing mechanisms: destruction of lung tissue known as emphysema and inflammation of airways, leading to thickened airway walls and narrowed airway lumen. Diagnosis of COPD is based on airflow limitation as determined with lung function tests, but CT adds to our understanding of disease mechanisms by for instance enabling lung density measurements, which gives insight into emphysema distribution within the lungs. The airways with their tree-like structure have so far been more challenging to analyse. Automated methods are indispensable as the visible airway tree in a CT scan can include several hundreds of individual branches. Automation of measurements is difficult, however, due to the airway's complicated structure, which varies in size and shape; biologically between subjects and dynamically during breathing.

This thesis presents several methods for solving problems related to analysing airways and results of using these methods to study COPD via data from the Danish Lung Cancer Screening Trial. A first step in analysing structures from images, is segmentation. A graph based approach for this is presented, which is able to simultaneously find the inner and outer airway wall surfaces in three dimensions. Measurements of airway wall surfaces depend on the branch they are made in. A method is presented that is able to match branches in multiple scans of the same subject using image registration, allowing measurements to be followed within single branches over time. Between subjects, measurements can be compared if they are done in the same anatomical subset of branches. A supervised algorithm for anatomical labelling of branches based on geodesic distances is presented, allowing such comparisons. Branch matching only allow changes from scan to scan to be measured if the branches are successfully segmented in each scan. To improve on this, the segmentation approach is extended to allow multiple scans to be segmented simultaneously by combining information from all images. The resulting surfaces correspond, meaning changes can be measured locally without the need for a separate branch matching step.

The developed fully automatic framework is applied to study effect of differences in inspiration level at the time of scan on airway dimensions in subjects with and without COPD. Results show measured airway dimensions, in scans close to maximum inspiration, to be affected by un-intended differences in the level of inspiration and this dependency is again influenced by COPD. Inspiration level should therefore be accounted for when measuring airways, and airway abnormalities typically associated with airflow limitation, such as airway wall thickening and lumen narrowing, should at least partly be understood as being due to differences in how airways are influenced by the inspiration level in subjects with and without COPD.

Resumé

Computertomografi (CT) tillader uobstruerede tredimensionelle billeder af det indre af skannede objekter. Dette kan bruges i medicinske anvendelser til for eksempel at diagnosticere eller studere sygdomme som Kronisk Obstruktiv Lungesygdom (KOL) med begrænsede omkostninger for patienterne. KOL er en betydelig årsag til død og invaliditet i hele verden. Den har indflydelse på, hvordan lungerne fungerer igennem to konkurrerende mekanismer: Ødelæggelse af lungevæv, kaldet emfysem; og betændelse i luftvejene, der fører til fortykkede luftvejsvægge og forsnævring af luftvejslumen. Diagnosen af KOL er baseret på begrænset lungefunktion, men CT øger forståelsen af sygdomsmekanismerne ved for eksempel at muliggøre målinger af lungetæthed, der giver indsigt i fordelingen af emfysem i lungerne. Luftvejene med deres trælignende struktur har indtil videre været svære at analysere. Automatiske målinger er uundværlige, da det synlige luftvejstræ kan indeholde flere hundrede individuelle grene. Automatisering af målingerne er dog svært pågrund af luftvejenes komplicerede struktur, der varierer både i størrelse og form; biologisk mellem personer og dynamisk under vejrtrækningen.

Denne afhandling præsenterer flere metoder til at løse problemer relateret til analyse af luftveje, samt resultater af at bruge disse metoder til at studere KOL via data fra det danske lungescreeningsprojekt. Et første skridt i analyse af stukturer i billeder, er segmentering. Til segmentering præsenteres en grafbaseret metode, der er i stand til samtidigt at finde den indre og ydre overflade af luftvejsvæggen i tre dimensioner. Efterfølgende målinger af luftvejsoverfladerne afhænger af, hvilken luftvejsgren de er lavet i. I den forbindelse præsenteres en metode, som er i stand til at matche grene i flere skanninger af den samme person ved brug af billedregistrering. Metoden gør det muligt at følge målinger i enkelte grene over tid. Målinger kan sammenlignes mellem personer, hvis de laves i en anatomisk ens delmængde af grenene. En algoritme fremlægges til anatomisk mærkning af grene, der gør sådanne sammenligninger mulige. Matchning af grene tillader kun ændringer fra skanning til skanning at blive målt, hvis de samme grene findes i hver skanning. For at forbedre dette udvides segmenteringsmetoden, så den kan segmentere flere skanninger samtidigt, ved at kombinere information fra alle billeder. De fundne overflader stemmer overens, således at ændringer kan måles lokalt uden behov for efterfølgende matchning af grene.

De udviklede automatiske værktøjer udnyttes til at studere effekten af ændringer i vejrtrækningsdybden ved skanning på de målte luftvejsdimensioner i personer med og uden KOL. Resultaterne viser at luftvejsdimensionerne, tæt ved maksimal inspiration, er påvirkede af utilsigtede forskelle i vejrtrækningsdybden og størrelsen af denne påvirkning ændres med KOL. Der skal derfor tages højde for vejrtrækningsdybden, når luftvejene måles og luftvejsabnormiteter, som normalt associeres med KOL, såsom fortykkelse af luftvejsvæggen og forsnævring af luftvejslumen, skyldes tildels forskelle i hvordan luftvejene påvirkes af vejrtrækningsdybden i personer med og uden KOL.

Acknowledgements

Three years time seems to have passed incredibly fast and although there have been times wishing I had a regular day job, I am grateful to have had the opportunity to do something that was this enjoyable, challenging, and meaningful to me.

I would especially like to thank my supervisor Marleen de Bruijne for her tireless and thorough guidance throughout this project and for giving me the freedom to explore my ideas; my co-supervisor Asger Dirksen for his enthusiasm and the medical and statistical expertise he provided the project with; and finally Mads Nielsen for his advices and motivation in getting me into medical image analysis in the first place.

Efforts by a number of other people have helped make this thesis what it is: Pechin Lo, Lauge Sørensen, and Vladlena Gorbunova's work as part of the earlier lung project; Aasa Feragen's statistics on trees and graphs; and insights on lungs, airways and pathology provided by Mathilde M. W. Wille and Laura H. Thomsen.

I also appreciate having had the chance to work at the Centre for Medical Image Computing at University College London with Marc Modat, Manuel J. Cardoso and Sébastien Ourselin. Besides giving me the opportunity to be part of an inspiring research environment, they also made an effort to get me out of the office every once in a while.

I would also like to thank AstraZeneca for funding the project, including Lars Nordenmark, Magnus Dahlbäck and Maria Gerhardsson, and Jesper J. H. Pedersen, the principal investigator on the Danish Lung Cancer Screening Trial, the primary source of data for this work.

I have enjoyed working in the Image Group, both former and current members and participants of the social events are to blame for this. There are too many to mention, but I would especially like to thank Sarah M. N. Abel, Katrine H. Jensen, Lars L. Râket, Mahdieh Khanmohamadi, and Oswin Krause for all the fun and sometimes odd conversations in the office.

Lastly, I would like to thank my family and girlfriend Sabrina for their support during these three busy years.

Contents

Chapter 1

Introduction

Advances in medical imaging have allowed detailed three-dimensional views of what is inside organs and bodyparts. The massive amount of data available from such sources means that in many cases automated analysis is needed to help clinicians diagnose and study diseases. This thesis focuses on a small subset of such methods, which are relevant in the study of a part of the human respiratory system, the airways. New image analysis methods are described, evaluated, and used to gain new insights into what effect Chronic Obstructive Pulmonary Disease (COPD) has on airways.

1.1 Anatomy of lungs and airways

The term airways is commonly used to describe the passage that air flows through, from nose or mouth and into the lungs during inspiration. The passage begins at the nasal and oral cavities and ends in the pulmonary alveoli, where gas exchange with blood occurs (Figure 1.1). In this thesis we will use the term airways to refer to the lower part of this passage, that is, the tree-like structure beginning at the trachea. The trachea bifurcates into the left and right main bronchi, which supply air to the left and right lungs. Here they bifurcate further, decreasing in size as they do so. The first generations of large airways are lined with cartilage and are called bronchi, these begin to be replaced by cartilage free small airways, called bronchioles, when the diameter is approximately 1 mm or less. After the 14th bifurcation on average the airways become increasingly lined with alveoli and at generation 23 on average they terminate in alveolar sacs (Weibel 2009).

The left and right lungs are separate structures located on opposite sides of the heart. Each lung is divided by fissures into lobes. There are three lobes in the right lung; the upper, middle and lower lobes and two lobes in the left lung; the upper and lower lobes. There is no middle lobe in the left lung, however the corresponding region is called the lingula. Each lobe is further divided into segments, up to 10 in each lung. Each segment is supplied with

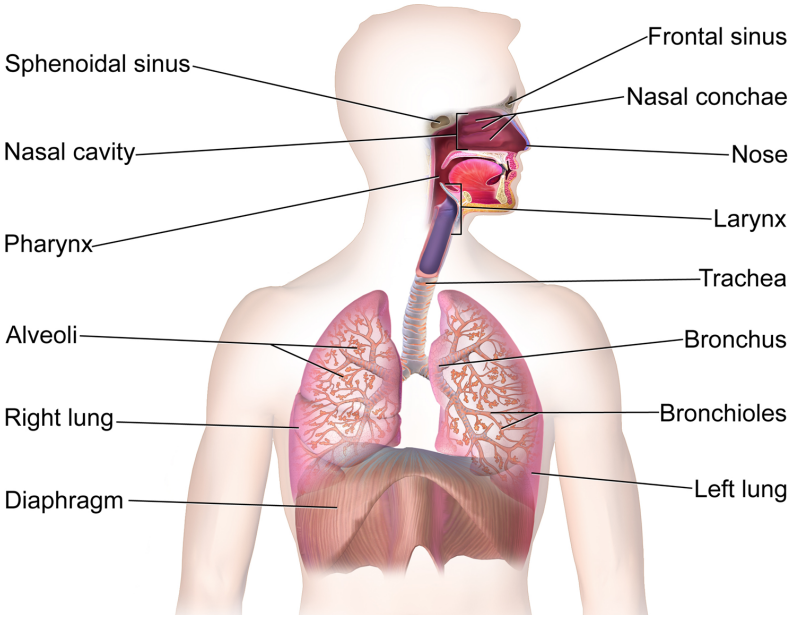


Figure 1.1: A schematic drawing of the human respiratory system with the tree-like structure of the airways: trachea, bronchi, bronchioles and alveoli (Image modified from Wikimedia Commons `http://commons.wikimedia.org/wiki/File:Blausen_0770_RespiratorySystem_02.png`)

air from one specific bronchus, known as a segmental bronchus (Chmura et al. 2008).

Airways are affected by diseases such as asthma (Lederlin et al. 2012), cystic fibrosis (Wielputz et al. 2013), and COPD (Hackx et al. 2012). In this thesis we will focus on abnormalities associated with COPD, however, the presented methods and some of the insights gained could be relevant in the study of other respiratory diseases as well.

1.2 Chronic obstructive pulmonary disease

COPD was estimated to have killed 3 mio. people in 2011, making it the fourth leading cause of death in that year according to the World Health Organization. The cause is most often smoking, but there are other contributing factors such as air pollution and genetics (GOLD 2013). COPD is an obstructive lung disease, which affects flow of air to and from the lungs by two competing factors; emphysema and small airway disease.

Emphysema is destruction and disappearance of lung tissue (parenchyma) due to inflammatory processes, which leaves the lungs filled with holes. These holes capture air that would otherwise be exhaled during expiration and that compress nearby normally functioning tissue. The changes decrease the elas-

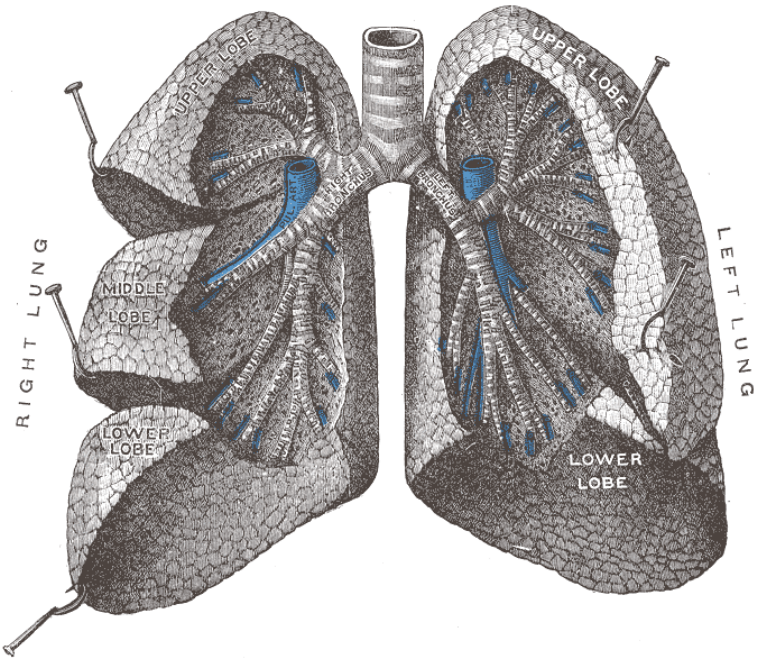


Figure 1.2: Right and left lungs with airways exposed. Lobes illustratively pulled apart. (Image from Wikimedia Commons `http://commons.wikimedia.org/wiki/File:Gray962.png`)

tic property, the so called elastic recoil, of the lungs making breathing more difficult (Hogg et al. 1969). Emphysema is also thought to affect lung function by destroying the airway's alveolar attachments. These attachments help keep the airways open during expiration when the pressure within the lung parenchyma is greater than inside the airways. Removal of the airway's alveolar attachments can thus lead to dynamic collapse of airways, which makes it difficult for subjects to expire (GOLD 2013).

Small airway disease is caused by inflammation within the airways, which causes structural changes such as airway wall thickening and airway narrowing. Increased mucus production may also cause airway plugging or increased airflow resistance (Hogg et al. 2004). It is called small airway disease because measurements have shown that airflow resistance is especially increased in peripheral and smaller airways compared to the central or larger airways (Hogg et al. 1968).

Diagnosis of COPD is based on lung function tests, also called spirometry. Two measures are critical; Forced Expiratory Volume in one second (FEV_1) and Forced Vital Capacity (FVC), which is the amount of air the subject can forcibly blow out after full inspiration in the first second and in total, respectively. Diagnosis of COPD is made if the ratio of FEV_1 to FVC is below 0.70. A predicted value of what FEV_1 should be if the subject had normal

GOLD group	Severity	Lung function
GOLD 1	Mild	FEV_1 % predicted $\geq$ 80 %
GOLD 2	Moderate	50% $\leq$ FEV_1 % predicted $<$ 80%
GOLD 3	Severe	30% $\leq$ FEV_1 % predicted $<$ 50%
GOLD 4	Very severe	FEV_1 % predicted $<$ 30%

Table 1.1: GOLD categories of airflow limitation according to GOLD 2013

lung function can be made based on gender, age and height (Pellegrino et al. 2005). Severity of airflow limitation in COPD subjects is then estimated based on FEV_1 as a percentage of its predicted value (FEV_1 % predicted) as shown in Table 1.1. The Global Initiative for Chronic Obstructive Lung Disease recently suggested to extend the classification of disease severity by combining assessments of airflow limitation with other symptoms (GOLD 2013). This new classification will not be used in this thesis.

Although diagnosis is based on lung function test, such tests provide little information on the underlying causes of airflow limitation. Computed Tomography (CT) images can be used to provide additional information, for instance to assess the roles of the two components of COPD: emphysema and small airway disease. This may be relevant in patient care or in studies to provide further insight into the disease (Hackx et al. 2012).

1.3 Quantitative analysis using computed tomography

Computed Tomography (CT) is used to construct virtual two-dimensional slices of three-dimensional objects, which allows un-occluded views of what is inside. Because of the huge advantage this has been to medical imaging, CT has now become almost synonymous with medical X-ray CT scans, which use X-rays as a source to create images. Modern so called multi-slice CT scanners, allow many slices to be collected at the same time and so can be used to create three-dimensional images of the body in a short amount of time (Figure 1.3).

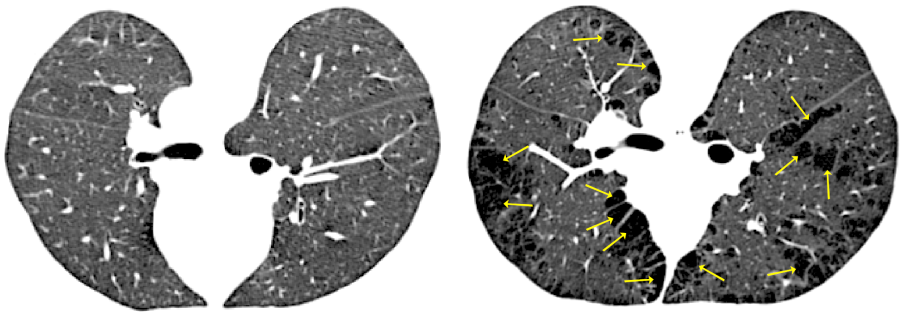
Image intensity in CT is a quantitative measure of radiodensity expressed on the Hounsfield Unit (HU) scale. On this scale -1000 HU corresponds to the radiodensity of air, whereas 0 HU corresponds to the radiodensity of water. The voxel values of the air-filled lung parenchyma will depend on the inspiration level, but typically the values range from -1000 HU to -700 HU, and the lungs therefore have a high contrast with the lung border, which has a density ranging from that of water (0 HU) to that of bone (typically around 400 HU to 1000 HU). Other structures within the lungs such as blood vessels and airway walls are closer to water in density and thus have values of around

Figure 1.3: Three-dimensional CT image of the chest region created by stacking individual CT slices.

0 HU. The high contrast between lung parenchyma and surroundings makes it relatively easy to extract the lungs using automatic methods.

The process of partitioning images into sub-structures or segments is called image segmentation. Automatic segmentation of lungs from CT images can be done using region growing methods (Lo 2010). Such methods work by iteratively extending the segmented region from an initial seed point by continuously adding nearby regions with density values of lung parenchyma (an upper threshold of -400 HU was used in (Lo 2010)) until the lungs are completely filled. Emphysema can be seen as darker regions within the lung, that is, regions with intensity closer to that of air than those of normal lung parenchyma. The standard way to quantify emphysema is using density measures, such as the percentage of the segmented lungs with intensity below -950 HU (Müller et al. 1988) or the 15^{th} percentile density (Gould et al. 1988). Figure 1.4 shows an example of a CT lung slice with normal and emphysematous lung parenchyma.

Airway abnormalities are much harder to quantify from CT images. Visually, signs of airway wall thickening and lumen narrowing are often too subtle to differentiate from normal variation in COPD cases (Figure 1.5). Quantification of airway abnormalities therefore needs accurate and precise



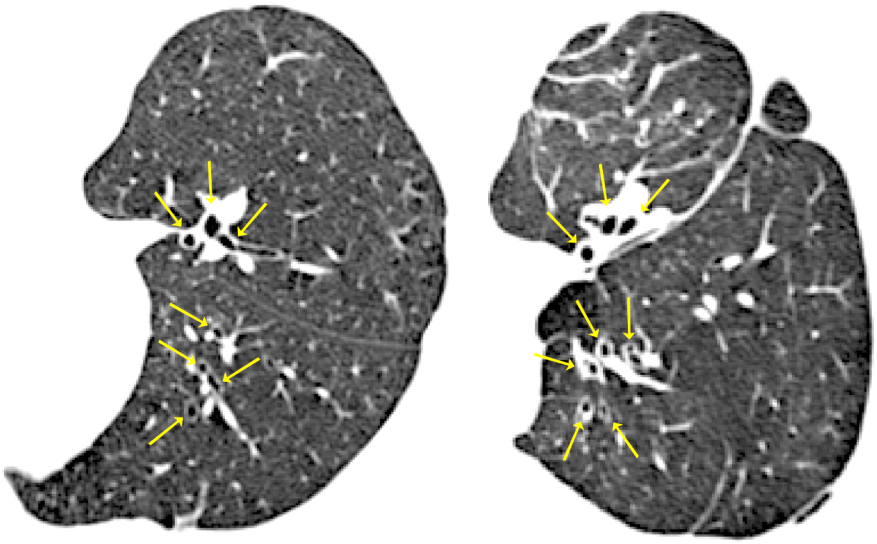
(a) Normal (b) Emphysema

Figure 1.4: An example of a CT lung slice with normal and emphysematous lung parenchyma. Arrows point to examples of large regions with emphysematous tissue, so-called bullae.

measurements, something which is simply too time consuming to manually do in all but a very limited number of locations in each scan. The visible airway tree in a CT scan may include several hundreds of individual branches, and automated methods thus allow a far greater amount of the available information to be exploited. The largest bronchioles are right at the resolution limit imposed by current clinical or low-dose CT images. This means that quantitative measurements cannot be consistently done with these images at the site in the airways where airflow resistance is most increased in COPD. However, it is thought that the same pathophysiological processes that lead to increased airflow resistance in small airways could be responsible for similar changes in large airways. Airway abnormalities in large airways, although perhaps having little influence on lung function, could thus still be used as indicators for small airway disease (Nakano et al. 2005).

In order to identify signs of COPD such as airway wall thickening, methods that can find the position of both the inner and outer airway wall surface are needed. This is typically realised as a two-step approach, first extraction of the airway tree (the inner or lumen surface) and second a step which uses the position of the airway tree to identify the position of the outer airway wall surface (Estépar et al. 2006; Weinheimer et al. 2008; Petersen et al. 2010; Saragaglia et al. 2006; Liu et al. 2012; Ortner et al. 2010).

Airway lumen in CT images appear as a connected and branching dark tubular structure bounded by airway wall, which appears as a lighter shell. Automatic extraction of this structure is a difficult task and an active area of research (Lo et al. 2012). Methods usually exploit the connected and bounded properties to iteratively grow the lumen from a seed point similar to how lung segmentation methods often work. Connectivity is, however, not necessarily given in case of pathologic airways, which may be collapsed, narrowed, or plugged with mucus, making such methods sensitive to disease severity.



(a) Normal (b) Small airway disease

Figure 1.5: CT lung slice showing example of normal and thickened airway walls. Arrows point to examples of visible airways.

Because airways become progressively smaller as they bifurcate, they also become progressively harder to see and the chance that airways are missed thus increases with each bifurcation. False positives in the form of spurious branches or leaks, in some cases including large parts of the parenchyma, may also be included in the extracted airway tree. This primarily happens when thickness of the airway wall approaches the image resolution, lung parenchyma may then locally appear, due to noise and partial volume effects, as if it is connected to the airway lumen, which causes the growing process to continue out of the airways.

The airway wall often abuts other nearby structures with similar radiodensity such as blood vessels and contrast between wall and lung parenchyma is also smaller than between wall and lumen. The outer wall surface is therefore more difficult to find than the inner lumen surface. A useful strategy is to exploit knowledge of the shape and position of the inner surface to help find the outer surface. This knowledge along with assumptions of tubularity of each airway branch have previously been used to search for the wall surfaces (Petersen et al. 2010; Weinheimer et al. 2008; Li et al. 2006). Such methods are, however, unable to segment the bifurcation regions where assumptions of tubularity break down. It is possible that newer methods, which allow these regions to be segmented (Ortner et al. 2010; Liu et al. 2012; Saragaglia et al. 2006 and Chapter 2), could provide independent information relevant in the study of diseases.

Additional challenges in measurements of airway dimensions are due to the airway tree's complicated structure, which varies biologically in topology, size and shape between subjects and dynamically in size and shape during breathing. It is therefore not straight-forward to conduct reproducible and comparable measurements of airway properties even given perfectly extracted airway trees. The dynamic variation of airways or the degree to which they expand and contract with the lungs during inspiration is called airway distensibility (Diaz et al. 2012; Brown et al. 2001). Airways are typically extracted from breathhold scans. It is, however, not easy for subjects to inspire to the same level in each scan. Distensible airways and un-intended variations in inspiration level may thus interact to create similar un-intended variations in airway measurements. The size of these effects is currently largely unknown.

1.4 Outline of thesis

This thesis consists of a series of chapters, which are either descriptions of methods solving parts of the problem of how to analyse airways from CT images or results of applying these methods to study the properties of airways. Each chapter is readable separately and is published or in preparation for being published. In addition, contributions to a number of other works were made during the period of this PhD. A full list of the resulting publications can found at the end of this thesis.

A first step in analysis of structures from images is segmentation. Chapter 2 presents an airway wall segmentation method, which is able to simultaneously find the inner and outer airway wall surfaces from CT images in three-dimensions across bifurcations. Chapter 3 presents a branch matching approach, which is able to match the segmented branches in multiple images of the same subject. It allows measurements of the same branches in different scans of the same subject to be compared, which removes some of the variability in measurements that occurs due to missing and spurious branches. Chapter 4 presents a branch labelling approach, which is able to identify the anatomical names of the airway branches down to and including the segmental level. It allows measurements of the same anatomical branches in different subjects to be compared, which removes some of the variability that occurs due to not measuring the same location within the airway trees. Chapter 5 and 6 apply the methods of Chapter 2, 3, and 4 to study the effect of differences in the level of inspiration at time of scan on measured airway dimensions in subjects with and without COPD from the Danish Lung Cancer Screening Trial (DLCST). The results show that differences in measured airway dimensions from multiple images of the same subject can be used to gain new insights on both longitudinal and dynamic changes related to COPD. This inspired the development of a joint approach to segmentation of airway wall surfaces in multiple images of the same subject, which is presented in Chapter 7. It

allows information from all images to be combined, which improves segmentation results. The method outputs corresponding surfaces, which make it possible to track differences from scan to scan locally in any point on the surfaces without the need for a separate branch matching step.

1.5 Main contributions

The main contributions of this thesis are:

- A new three-dimensional segmentation method, which can segment both airway wall surfaces simultaneously including bifurcations regions. It can use the position of one surface to help position the other and it is designed to handle high curvature regions, such as the bifurcation areas of the airway tree better than similar methods from the literature. The method is validated using manual annotations, a phantom scan, and expert visual inspection. Airway abnormality measurements obtained using the approach are reproducible and correlate significantly with lung function (Chapter 2).

- Branch matching allows measurements of airway dimensions in individual branches to be compared in multiple scans of the same person. An approach which uses deformable image registration to match branches is presented. Results show increased reproducibility if measurements are limited to repeatedly found branches, indicating the method could be useful to study changes over time or due to differences in inspiration level (Chapter 3).

- A new supervised algorithm for anatomical labelling of airway branches based on tree-space geodesics, which performs similar to medical experts in terms of accuracy, better in terms of reproducibility, and robustly with respect to COPD. The detailed evaluation provides insights into not only performance of the algorithm but also that of medical experts (Chapter 4).

- A study on the effects of inspiration level differences in maximum inspiration scans on measurements of airways in subjects with and without COPD. A model of the effects is provided and a way to adjust measurements for them is suggested. Results show how airway dimensions and distensibility vary with airway position and COPD severity (Chapter 5 and 6).

- An extension of the airway wall segmentation method of Chapter 2 to handle multiple images jointly using groupwise image registration. The approach enables information from all the images to be combined to improve segmentation results and it outputs corresponding inner and

outer airway wall surfaces for all images, allowing differences to measured locally at any point on them without the need for a separate branch matching step. The method has the potential to improve detection of longitudinal and dynamic changes using images of different time-points and inspiration-levels (Chapter 7).

Chapter 2

Optimal surface segmentation using flow lines

The work presented in this chapter is based on J. Petersen, M. Nielsen, P. Lo, L. H. Nordenmark, J. H. Pedersen, M. M. W. Wille, A. Dirksen, and M. de Bruijne (2014). "Optimal surface segmentation using flow lines to quantify airway abnormalties in chronic obstructive pulmonary disease". In: *Med Image Anal* (In Press). URL: `http://dx.doi.org/10.1016/j.media.2014.02.004`.

Abstract

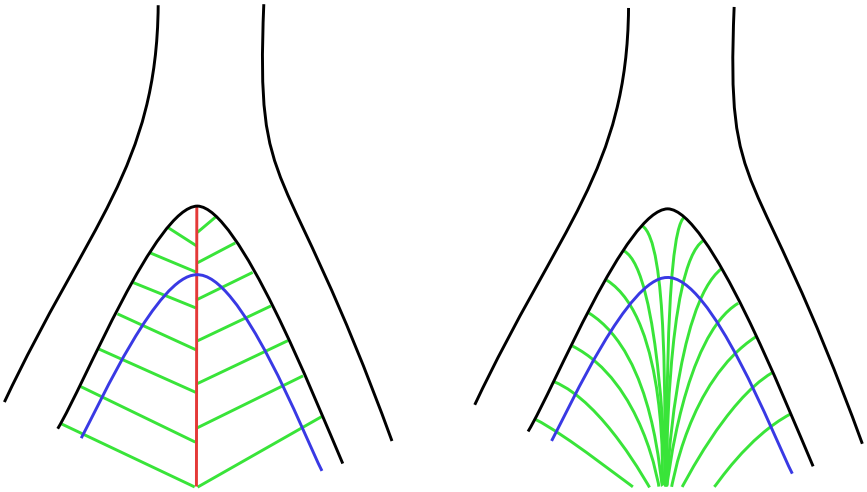
This chapter introduces a graph construction method for multi-dimensional and multi-surface segmentation problems. Such problems can be solved by searching for the optimal separating surfaces given the space of graph columns defined by an initial coarse surface. Conventional straight graph columns are not well suited for surfaces with high curvature, we therefore propose to derive columns from properly generated, non-intersecting flow lines. This guarantees solutions that do not self-intersect.

The method is applied to segment human airway walls in computed tomography images in three-dimensions. Phantom measurements show that the inner and outer radii are estimated with sub-voxel accuracy. Two-dimensional manually annotated cross-sectional images were used to compare the results with those of another recently published graph based method. The proposed approach had an average overlap of 89.3 ± 5.8 %, and was on average within 0.096 ± 0.097 mm of the manually annotated surfaces, which is significantly better than what the previously published approach achieved. A medical expert visually evaluated 499 randomly extracted cross-sectional images from 499 scans and preferred the proposed approach in 68.5 %, the alternative approach in 11.2 %, and in 20.3 % no method was favored. Airway abnormality measurements obtained with the method on 490 scan pairs from a lung cancer screening trial correlate significantly with lung function and are reproducible; repeat scan R^2 of measures of the airway lumen diameter and wall area percentage in the airways from generation 0 (trachea) to 5 range from 0.96 to 0.73.

2.1 Introduction

Optimal net surface methods (Wu and Chen 2002) have seen a growing use within medical image segmentation in the last couple of years, likely due to their ability to find the globally optimal solution of multiple interacting surfaces in multiple dimensions given surface cost functions and a useful range of geometric constraints and penalties in polynomial time using minimum cut algorithms (Li et al. 2006; Liu et al. 2012; Petersen et al. 2010; Yin et al. 2009; Abràmoff et al. 2008; Petersen et al. 2011b; Arias et al. 2012; Kainmueller et al. 2013). In order to use these methods, the segmentation problem needs to be transformed from the space defined by the image voxel grid to some graph representation defined by a set of columns. Each column is associated with a point on the sought surface and represents the set of possible solutions, or positions, the surface can take. A suitable graph should be able to represent all plausible solutions in the image space. For instance, if a graph column does not cross the sought surface or if it crosses it multiple times, then this surface can not be represented by the graph. Similarly, admissible solutions in the space defined by the graph representation should represent valid surfaces in image space, that is, the graph space should for instance not allow self-intersecting surfaces. It is also important that the graph structure allows for a meaningful representation of the surface cost functions and geometric constraints and penalties. Surface non-smoothness can, for instance be reduced, by increasing the cost of solutions in proportion to how much they vary in neighbouring columns. However this is only meaningful if the relative variation within the columns is somehow related to the associated relative variation within image space.

In some cases the sought surfaces are expected to be oriented along an image axis and the voxel columns of the image itself may be used. This has for instance been used in the case of the intraretinal layers in macular optical coherence tomography images (Abràmoff et al. 2008). Other approaches have used simple mathematical transformations, such as those of Li et al. 2006 and Petersen et al. 2010, in which images of tubular airway segments were unfolded using polar transforms in two or three-dimensions. The graph columns were oriented perpendicular to the resulting contours or terrain like surfaces allowing for an easy representation of surface smoothness constraints and penalties. In many cases, however, the surfaces are much too complicated for such an approach and/or the prior knowledge of the surfaces' shape and position required is not available. In these cases such prior knowledge may be gained by employing an initial method to roughly estimate the position of the surfaces and then use an optimal surface graph to refine this estimate. This was done in Liu et al. 2012 by placing columns at points of the initial surface and oriented along the surface normals inward and outward. Problems with intersecting columns and thus self-intersecting surfaces were avoided, by limiting the length of each column to the minimum distance to the initial surface



(a) Medial Axes (b) Flow Lines

Figure 2.1: Figure 2.1a illustrates the fish-bone like structure of surface normal direction columns (*green*) based on the distance to the medial axis (*red*) in areas where the initial segmentation (*black*) has high curvature. Notice that the four inner-most columns do not cross the sought surface border (*blue*), which means that the desired solution can not be represented by the graph and the segmented surface will be wrong in these positions. Figure 2.1b shows the advantage of columns based on flow lines (*green*), notice that all columns cross the sought surface.

inner and outer medial axes. This approach can result in columns that are too short to reach the desired solution, as shown in Figure 2.1a. Yin et al. 2009 suggested columns inspired by the non-intersecting property of electric lines of force. The columns were constructed by simulating electrical charges at surface points of the initial segmentation and tracing the electric lines of force within the field inward and outward. This method is computationally infeasible for large scale problems, as every surface point charge influences the computation of every electric line of force. Furthermore the electric lines of force can behave erratically if the initial segmentation contains small scale errors or noise. Recently Kainmueller et al. 2013 proposed to use omnidirectional displacements, which allow each initial surface mesh vertex to move to uniformly distributed positions within a ball shaped region around it. Self-intersections are minimized by using regularisation and the solution is found using Markov Random Field energy minimization. The approach is too computationally expensive for larger problems and so Kainmueller et al. 2013 also shows how it can be combined with an optimal surface - unidirectional column type approach. This makes the method practical for larger problems by using omnidirectional displacements in high curvature regions and unidirectional

columns in low curvature regions. The two problems are solved sequentially and so the method does not guarantee global optimality.

In Petersen et al. 2011b we proposed to use graph columns defined from flow lines within a regularized version of the initial segmentation. Flow lines are non-intersecting and are uniquely defined if the regularisation is smooth, and noise and small errors in the segmentation are naturally dealt with by the same regularisation. Moreover, fast approximations can be computed using image convolution. Figure 2.1b illustrates the concept. The method was originally applied to the problem of segmenting human airway walls in CT images and has since then been used for segmenting the carotid artery bifurcation in magnetic resonance imaging (Arias et al. 2012).

Assessing the dimensions of the airway walls is important in the study of airway remodelling diseases such as Chronic Obstructive Pulmonary Disease (COPD) (Hackx et al. 2012). It is a dual surface problem, consisting of an inner and an outer wall surface, where bifurcations form regions of high curvature that would cause problems for conventional graph construction approaches. The vast majority of previous airway wall segmentation methods have been one- or two-dimensional in nature. The one-dimensional techniques work by casting rays from the centre of the airways outwards looking for the wall surfaces using the full width at half maximum edge detection principle (Nakano et al. 2000), by phase congruency (Estépar et al. 2006), or more complex models of the scanning point spread function (Weinheimer et al. 2008). The airway wall surfaces resemble concentric circles when seen in a cross-sectional view centred on and perpendicular to the airway centreline. This is what two-dimensional methods typically exploit to impose some degree of regularity on the solution (Petersen et al. 2010; Saragaglia et al. 2006). Three-dimensional methods, however, may use more of the information present in the image, allowing surfaces to be found more accurately when they are close to other structures such as blood vessels. Moreover, bifurcation and carina regions, which typically cannot be segmented with previous two-dimensional approaches, can be analysed (Liu et al. 2012). Besides the already mentioned methods of Liu et al. 2012; Petersen et al. 2011b, a three-dimensional method is also described in Saragaglia et al. 2006, which evolves a deformable mesh, constructed from an initial segmentation of the lumen. The evolution is done with force constraints computed from intensity and gradient magnitude values; elastic forces penalizing local wall thickness variations; and regularisation forces, locally smoothing the result. The method does not guarantee a global optimal solution and unlike the approaches of Liu et al. 2012; Petersen et al. 2011b the two surfaces are not estimated simultaneously, and thus the added knowledge of the position of the exterior surface is not used to improve the inner surface. Ortner et al. 2010 also proposed to use a deformable mesh. Their mesh is built from an initial segmentation of the lumen and its evolution is governed by gradient vector flow and simplified Lagrangian dynamics and so avoids self-intersections. The approach was evaluated on simulated CT data

and 15 clinical cases of mild and severe asthmatics, showing good agreement with segmentation result and clinical expertise.

This chapter is an extension of the work presented in Petersen et al. 2011b. The main differences are the addition of a constraint, that forces the outer surface to be outside the inner; improvements in the parameter tuning, such that all involved parameters are automatically estimated using a manually annotated training set; adjustment of parameters and evaluation of results according to the COPDGene phantom (Sieren et al. 2012) to account for a possible bias present in the manual annotations; and finally the addition of an extensive medical expert visual evaluation comparing the result of the proposed approach with that of Liu et al. 2012. We show that the method can be used to measure airway abnormalities associated with COPD reproducibly.

2.2 Methods

2.2.1 Initial Segmentation

We will assume the existence of a coarse initial segmentation, a single object given by the voxels in the set S, whose surface should be roughly similar to the surfaces we are looking for. For our application we used an airway tree extraction algorithm based on Lo et al. 2009, which returns a three-dimensional binary segmentation of the airway lumen.

The initial segmentation needs to be converted to a mesh. To this end, we used vertices at the centre of each surface voxel face and the neighbourhood given by the face edge neighbours, such that each vertex has 4 neighbours. We will denote the vertices in this mesh with $\mathcal{V}$, and represent the neighbourhood with an edge set $\mathcal{E}$, where $(i, j) \in \mathcal{E}$ denotes that the vertices $i, j \in \mathcal{V}$ are neighbours.

2.2.2 Flow lines

The graph will be defined from flow lines at each of the mesh vertices. A flow line in a vector field is tangent to the field at each point, and if the field is defined as the gradient field of a scalar potential ϕ, it will follow the direction with the greatest rate of change of this potential. Electric lines of force are examples of flow lines in an electric field and interestingly the magnitude of the electric field can be formulated as a convolution operation. The approach of Yin et al. 2009 can thus be efficiently approximated using fast convolution algorithms. Rather than being limited to electric fields, we propose, however, to use the more general case of flow lines based on scalar potentials defined by the convolution:

$$\phi(\boldsymbol{x}) = \int \mathbf{1}_S(\hat{\boldsymbol{x}}) R(\hat{\boldsymbol{x}} - \boldsymbol{x}) d\hat{\boldsymbol{x}} \ , \tag{2.1}$$

here $\boldsymbol{x}$ is the position to be evaluated and $\mathbf{1}_S$ is the indicator function for the initial segmentation S.

In this work we experimented with generating the potentials from two different types of filters. The first can be considered a regularised electric line of force approach:

$$R(\boldsymbol{x}) = \frac{1}{\alpha + |\boldsymbol{x}|^2} \ , \tag{2.2}$$

where $\alpha > 0$ is the regularisation constant, which makes R well defined for all $\boldsymbol{x}$. This regularisation has the added effect of smoothing the result, which is useful if the initial segmentation contains noise. When $\alpha \to 0$, ϕ becomes proportional to the magnitude of the electric field at x arising from a continuous 'charge density' given by $\mathbf{1}_S$. This option is therefore similar to the method introduced in Yin et al. 2009. However, rather than using a discrete set of surface point 'charges', which introduce local singularities, ϕ is defined everywhere and thus allow us to trace the flow lines consistently through the initial surface. The second filter is given by the Gaussian kernel, which offers more regularisation and more locality because of its faster decaying tails:

$$R(\boldsymbol{x}) = ce^{-|\boldsymbol{x}|^2/(2\sigma^2)}. \tag{2.3}$$

The flow lines $\varphi_i : \mathbb{R} \to \mathbb{R}^n, i \in \mathcal{V}$ are found as the solution to the following ordinary differential equation:

$$\frac{\partial \varphi_i}{\partial t}(t) = \nabla\phi(\varphi_i(t)) \tag{2.4}$$

with initial value $\varphi_i(0) = i$. An implementation (*GNU Scientific Library Reference Manual (3rd Ed.)*) of the Runge-Kutta-Fehlberg method (Fehlberg 1970) was used to approximate the solutions. At some point the gradient flattens such that the flow line can no longer be traced due to numerical issues or the limited support of the employed discrete convolution kernel, resulting in a column with a finite number of inner and outer column points relative to i.

With no approximations the running time of computing electric lines of force as suggested in Yin et al. 2009 is given by $O(\eta \times |V|^2)$, where η is the number of times the gradient needs to be computed to trace each electric line of force. This is because the charge at each of the mesh vertices influences the computation of the gradient needed to compute the electric line of force at each of the other mesh vertices. However using discrete convolution, an approximated gradient can be computed efficiently and represented as an image. Computing the gradient in an arbitrary position is then a matter of simple interpolation, and the running time of computing flow lines as suggested in this work is thus $O(\eta \times |V|)$.

2.2.3 Graph construction

This section describes how an optimal surface graph $G = (V, E)$ with vertices V and edges E can be constructed, such that the minimum cut of G defines the sought surfaces.

We will use the term *penalty* to describe what could be called a soft constraint, that is, something that has the effect of increasing the cost and decreasing the likelihood of a given solution. A *constraint* on the other hand refers to a condition the solution is required to satisfy.

The vertices V of the graph G are arranged in columns $V_i^m = \{i_k^m \mid k \in K_i\}$, one for each vertex $i \in \mathcal{V}$ of the initial surface mesh and for each sought surface $m \in M$, plus source s and sink t vertices. $K_i = \{-I_i, 1-I_i, ..., 0, ..., O_i\}$ denote the indices associated with the sampled flow line with I_i and O_i inner and outer column points relative to i. In this way, the columns associated with each sought surface are corresponding, that is, they represent the same set of possible positions, given by the sampled flow lines. We therefore have:

$$V = \bigcup_{i \in \mathcal{V}, m \in M} V_i^m \cup \{s, t\} \ . \tag{2.5}$$

In the case of airway wall segmentation, $M = \{0, 1\}$ would denote the fact that there is an inner and outer surface sub-graph.

Let $w_i^m(k) \geq 0$ denote the surface cost function or the data term of the optimization problem, which maps a vertex with index $k \in K_i$ in a column V_i^m to the cost of making this vertex part of the surface m, see Section 2.2.4. Also let $f_{i^m,j^n}(|k-l|)$ be a convex non-decreasing function describing the pairwise cost (or penalty) of vertices $i_k^m \in V_i^m$ and $j_l^n \in V_j^n$ being part of the solution of the surfaces $m, n \in M$ respectively. These pairwise penalties are used to implement surface smoothness and separation penalties, see Equation (2.8). The vertices defining the sought surfaces $\mathcal{N} \subseteq V$, are then a solution to the following minimization problem:

$$\begin{aligned} \mathcal{N} = \underset{\hat{\mathcal{N}} \subseteq V}{\arg\min} & \sum_{i_k^m \in \hat{\mathcal{N}}} w_i^m(k) + \sum_{i_k^m, i_l^n \in \hat{\mathcal{N}}} f_{i^m,j^n}(|k-l|) \\ & \text{s.t. } \exists! k (i_k^m \in \hat{\mathcal{N}}) \\ & \quad i_k^m, j_l^n \in \hat{\mathcal{N}} \Rightarrow l \in \Omega(i_k^m, j^n), \end{aligned} \tag{2.6}$$

The first of the constraints ($\exists! k(i_k^m \in \hat{\mathcal{N}})$) ensures that one and only one vertex in each column is part of the solution. This is a needed requirement for the method to work, and it also has the effect that the topology of the initial segmentation is preserved in the solution surfaces. The second constraint ($i_k^m, j_l^n \in \hat{\mathcal{N}} \Rightarrow l \in \Omega(i_k^m, j^n)$) enforces pairwise limits on which vertices can be included, by using the set $\Omega(i_k^m, j^n) = \{l', l'+1, ..., l'+\delta\} \subseteq \{-I_j, 1-I_j, ...O_j\}$. That is, if some vertex i_k^m is part of the solution $\mathcal{N}$, then the solution must include one of the vertices in $j_{l'}^n, j_{l'+1}^n, ..., j_{l'+\delta}^n$ as well.

In our experiments we define the data term based on image derivatives, as explained in detail in Section 2.2.4. We used the following edge constraints to force the outer surface to be outside the inner:

$$\Omega(i_k^m, j^n) = \begin{cases} \{k, k+1, ..., O_j\} & \text{if } m = 0 \text{ and } n = 1 \\ K_j & \text{otherwise.} \end{cases} \tag{2.7}$$

The following pairwise penalty functions were implemented:

$$f_{i^m,j^n}(x) = \begin{cases} p_m x & \text{if } m = n \text{ and } (i,j) \in \mathcal{E} \\ qx & \text{if } m \neq n \text{ and } i = j \\ 0 & \text{otherwise,} \end{cases} \tag{2.8}$$

where $\mathcal{E}$ is the neighbourhood defined in Section 2.2.1, p_m is the smoothness penalty, defining the cost of each index the solution varies between neighbouring columns in the same surface m, and q is the separation penalty, defining the cost for each index the surfaces are separated in each column.

Next we will describe how to construct the edge set E, such that the solution given by a minimum s-t cut in G satisfies Equation 2.6. It should be noted that our construction differs from previous methods (Ishikawa 2003; Wu and Chen 2002), in its ability to deal with columns of varying inner and outer length. First we will note that the edges of E are directed and are associated with a capacity and we will use the following notation $(v \xrightarrow{c} u)$, to indicate an edge from vertex v to vertex u with capacity c. We will also remind the reader that an s-t cut in a graph is a partition of the vertices of the graph into two disjoint subsets such that s and t belong to different subsets. The cost of the s-t cut, is the sum of the capacities of the edges going from the source-set (the specific subset s belongs to) to the sink set (the specific subset t belongs to). The vertices in the sought surfaces $\mathcal{N}$ will be given by the vertex $i_k^m \in V$ in each column, which has the highest index k and is part of the source-set.

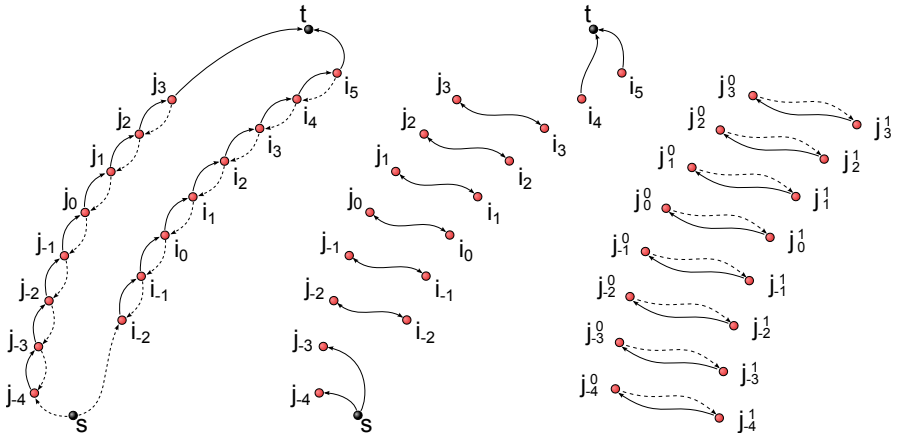
The data term can be implemented with the following edges:

$$\begin{aligned} E_d = &\left\{ (i_k^m \xrightarrow{w_i^m(k)} i_{k+1}^m) \mid i_k^m, i_{k+1}^m \in V \right\} \cup \\ &\left\{ (i_{O_i}^m \xrightarrow{w_i^m(O_i)} t), (s \xrightarrow{\infty} i_{I_i}^m) \mid i_{O_i}^m, i_{I_i}^m \in V \right\}. \end{aligned} \tag{2.9}$$

Since each column is a direct line of flow from the source to the sink, it will always be cut at least once. However in some degenerate cases, multiple cuts might exist in each column violating the first constraint of Equation 2.6. To prevent this, infinite cost edges directed opposite to the data term edges are added:

$$E_\infty = \left\{ (i_k^m \xrightarrow{\infty} i_{k-1}^m) \mid \ i_{k-1}^m, i_k^m \in V \right\}. \tag{2.10}$$

An example of these edges is given in Figure 2.2a.



(a) Data and topology (b) Smoothness (c) Separation

Figure 2.2: Two neighbouring columns V_i and V_j showing the graph construction as implemented in this chapter. These columns have $I_i = 2$ and $I_j = 4$ inner column vertices and $O_i = 5$ and $O_j = 3$ outer column points (note m subscript left out for clarity in 2.2a and 2.2b). The dotted edges have infinite capacity and implement the topology (2.2a) and separation constraints (2.2c). The solid edges have capacities as determined by the data term (2.2a), smoothness (2.2b) and separation penalty (2.2c). Here $L(j,i) = \{-4,-3\}$, $U(i,j) = \{4,5\}$, and $L(i,j) = U(j,i) = \{\}$.

The edge penalties and constraints can be implemented by the following edges:

$$\begin{aligned} E_i = \Big\{ & \big\{ (i_k^m \overset{\triangle(i_k^m, j_l^n)}{\rightarrow} j_l^n) \mid k \in K_i, l \in K_j \big\} \cup \\ & \big\{ (s \overset{\triangle(i_k^m, j_l^n)}{\rightarrow} j_l^n) \mid l \in K_j, k \in L(j,i) \big\} \cup \\ & \big\{ (i_k^m \overset{\triangle(i_k^m, j_l^n)}{\rightarrow} t) \mid k \in K_i, l \in U(i,j) \big\} \\ & \mid i,j \in \mathcal{V}, m,n \in M \Big\} , \end{aligned} \tag{2.11}$$

where L and U are the needed lower and upper edge endpoints, which are missing due to differences in column inner and outer lengths, see Figure 2.2 for an example:

$$\begin{aligned} L(i,j) = & \quad \{k \mid k \in K_i, k < -I_j\} \\ U(i,j) = & \quad \{k \mid k \in K_i, k > O_j\}, \end{aligned} \tag{2.12}$$

and $\triangle$ is the capacity of the edges calculated from the pairwise penalty func-

tion:

$$\triangle(i_k^m, j_l^n) = \begin{cases} \infty & \text{if } l = \min \Omega(i_k^m, j^n) \\ 0 & \text{if } l \notin \Omega(i_k^m, j^n) \\ \hat{\triangle}_{i^m,j^n}(k-l) & \text{otherwise} \end{cases} \tag{2.13}$$

where

$$\hat{\triangle}_{i^m,j^n}(x) = \begin{cases} 0 & \text{if } x < 0 \\ f_{i^m,j^n}(1) - f_{i^m,j^n}(0) & \text{if } x = 0 \\ f_{i^m,j^n}(x+1) - 2f_{i^m,j^n}(x) + & \text{if } x > 0, \\ f_{i^m,j^n}(x-1) & \end{cases} \tag{2.14}$$

ensures that any solution will include the cost of the pairwise penalties. For a proof of why this is so refer to Wu and Chen 2002. Notice that $\triangle(i_k^m, j_l^n) \geq 0$ for all $i_k^m, j_l^n \in V$, so the edge capacities are all positive.

Note there is a computational advantage of using the linear pairwise penalty functions of Equation 2.8 compared to non-linear pairwise penalty functions, as far fewer edges are needed to implement them because $\hat{\triangle}_{i^m,j^n}(x) = 0$ for all $x \neq 0$. An illustration of these edges is given in Figure 2.2b and 2.2c.

The total edge set E in the maximum flow graph is given by:

$$E = E_d \cup E_\infty \cup E_i \ . \tag{2.15}$$

We used the algorithm described in Boykov and Kolmogorov 2004 to find the minimum cut.

2.2.4 Data term

In this section we describe how the data term part of Equation 2.9 was implemented for the airway wall segmentation problem. The columns, in this case, will usually start inside the air-filled lumen area, which has low density, move through the airway wall where the density rises, and finally end up in the lung parenchyma where the density falls again. The CT intensity directly reflects this density change. A common way to find such boundaries is to use weightings of the first and second order derivatives of the intensity along the columns (Li et al. 2006; Liu et al. 2012; Petersen et al. 2010). Because the extrema of the second order derivative are slightly shifted and on opposite sides of the first order derivative weighting the derivatives allows one to adjust the position of the found surface according to some known groundtruth. In this way one can adjust for bias introduced by partial volume effects:

$$\hat{w}_i^0(t) = \begin{cases} (1 - |\gamma_0|)d_i(t) + \gamma_0 \frac{\partial d_i(t)}{\partial t}(t) & \text{if } d_i(t) > 0 \\ 0 & \text{otherwise} \end{cases}, \tag{2.16}$$

$$\hat{w}_i^1(t) = \begin{cases} (|\gamma_1| - 1)d_i(t) + \gamma_1 \frac{\partial d_i(t)}{\partial t}(t) & \text{if } d_i(t) < 0 \\ 0 & \text{otherwise} \end{cases}, \tag{2.17}$$

where $\gamma_0, \gamma_1 \in [-1, 1]$ are the weights, $d_i(t) = \frac{\partial I \circ \varphi_i}{\partial t}(t)$, is the first order derivative of the image intensity I along the flow line φ_i. To get the actual cost functions, the functions are inverted:

$$w_i^m(k) = \max_{j \in \mathcal{V}, l \in K_j} \hat{w}_j^m(j_l^m) - \hat{w}_i^m(i_k^m). \tag{2.18}$$

Numerically we approximate the derivatives using central differences from cubically interpolated values.

2.3 Experiments

Experiments were conducted with three methods based on different ways of constructing the columns. One was a method using straight columns $S_{k,\tau}$, as described in Liu et al. 2012, where the medial axes and normals were determined using the method of Dey and Sun 2006 using k neighbours and an error tolerance of τ respectively. We refer to the original article for a definition of these parameters. The other two methods used the proposed flow line columns calculated using convolution kernels based on Equation 2.2 and using a Gaussian of scale σ, denoted F_α and F_σ respectively. The resolution of the initial mesh used in the experiments was 0.5 mm $\times$ 0.5 mm $\times$ 0.5 mm and the flow lines were sampled at 0.5 mm arc length intervals.

2.3.1 Data

The data comes from the Danish Lung Cancer Screening Trial (DLCST) (Pedersen et al. 2009). The images were obtained using a Multi Detector CT scanner (16 rows Philips Mx 8000) with a low dose (120 kV and 40 mAs), and reconstructed using a hard kernel (D) with a resolution of approximately 0.78 mm $\times$ 0.78 mm $\times$ 1 mm.

Lung function measurements were performed according to recommendations by the European Respiratory Society (Miller et al. 2005) using a computerized system (Spirotrac IV; Vitalograph, Buckingham, UK). No bronchodilation was applied.

As in Petersen et al. 2011b we use a randomly selected training and test set of 8 and 7 images for parameter tuning and evaluation. From the test and training data set we extracted 329 and 319 two-dimensional cross-sectional sub-images with a resolution of 0.5 mm $\times$ 0.5 mm at random positions perpendicular to and centred on the airways. We then manually annotated these images with lumen M_l and complete airway M_a area. Some of the sub-images contain regions, which we will denote X, which belong to other airway branches. As the orientation of these other branches is unknown annotation may be difficult, and they were therefore marked and excluded from the analysis. The COPDGene phantom (Sieren et al. 2012) was scanned with the DLCST protocol and the 6 airway-like tubes present in the phantom were

used to further adjust method parameters and validate segmentation accuracy. Additionally we randomly selected 499 subjects for the medical expert visual comparison. 490 of these had repeated scans and spirometry within a two year period and were further selected to evaluate reproducibility of measures of airway morphology as well as their correlation with lung function. Of these, 266 are men and 270 were found to be asymptomatic at first scan time, 143, 71 and 6 had COPD stage 1, 2 and 3 respectively according to the criteria of the global initiative for chronic obstructive lung disease (GOLD 2013). The average scan interval was 432 ± 32 days. The average absolute interval between scan and lung function measurement was 12 days.

2.3.2 Parameter tuning

The methods have inner and outer smoothness penalties, inner and outer cost function derivative weightings and separation penalties, denoted p_m, γ_m and q where $m \in \{0, 1\}$ respectively, plus the method specific parameters, that is k, τ (Section 2.3) and α and σ (Section 2.2.2). Optimal values of all these parameters were obtained by searching the parameter space on the training data set using an iterative algorithm. In each iteration, a parameter is searched by probing upper and lower search interval limits. If no improvement is detected in the average value of the error metric Φ, see Section 2.3.3, over all subjects in the training set, the intervals are halved around the current best guess and the process is repeated. This continues until a better guess is found or the difference between the upper and lower search intervals becomes less than a threshold. The search then proceeds with the next parameter in a loop with reset upper and lower search interval limits until no more parameters get updated. In order to avoid getting stuck in a local minimum, the algorithm was repeated 10 times using random values as initial guesses for the parameters.

2.3.3 Comparison with manual annotations

We evaluate the correctness of a segmentation in the cross-sections using the relative area of overlap outside the excluded area, Φ as follows:

$$\Phi(M_l, M_a, A_l, A_a, X) = \begin{array}{l} \frac{|(M_l \cap A_l) \backslash X|}{|M_l \backslash X| + |A_l \backslash X|} + \\ \frac{|(M_a \cap A_a) \backslash X|}{|M_a \backslash X| + |A_a \backslash X|} \end{array}, \tag{2.19}$$

where A_l and A_a denote lumen and complete airway area as found by the algorithms.

Define the contour $\mathcal{C}(A)$ of an area A as the set of pixel centres belonging to A, where at least one of the pixels in the standard 4-neighbourhood is not

part of A. The average contour distance was then defined by:

$$\Psi(M_l, M_a, A_l, A_a, X) = \underset{x \in \mathcal{C}(A_l \setminus X)}{\text{mean}} d(\mathcal{C}(M_l \setminus X), x)/2 + \underset{x \in \mathcal{C}(A_a \setminus X)}{\text{mean}} d(\mathcal{C}(M_a \setminus X), x)/2 \ , \tag{2.20}$$

where $d(A, x)$ defines the minimum euclidean distance between the point x and the set A. Note, that this validation is performed in resampled images of the resolution of the graph, which is higher than the original image resolution.

2.3.4 Phantom experiments

It has been established that humans tend to underestimate the lumen and overestimate the wall area (King et al. 2000) and graph segmentations tuned to manual segmentations are likely to be biased in the same manner. We therefore conducted a second round of tuning of the data term parameters γ_0 and γ_1 of all the methods involving the COPDGene phantom as ground truth. The airway-like tubes of the phantom were initially segmented using a region growing approach with an upper threshold of -900 HU. The unsigned relative deviation of the inner and outer radii on each tube was used as an error metric both to tune the parameters and to evaluate segmentation performance.

2.3.5 Observer study

As a final comparison between F_σ and $S_{k,\tau}$ we let a medical expert (MMWW) judge the quality of the segmentations using the phantom tuned parameters. For each subject in the data set a single random position in the airway tree was selected. In this position three cross-sectional images were extracted perpendicular to the centreline. One image containing the original scan data and the other two the original scan data overlaid with the segmentation results of the two methods. We automatically removed all segmented components not connected to the centre-most pixel, to restrict evaluation to the part of the airway viewed perpendicularly. The medical expert was presented with these three images in one view and was asked to decide among four options: 1) result of method a is best, 2) result of method b is best, 3) both results are of equal quality, or 4) not enough information is present to make the decision. The expert could scroll through the scans of the data at will, but was blinded to which method created them. When moving to the next scan, the position of the segmented images in the view would switch randomly. Figure 2.5 column 1, 3 and 4 illustrate how these images look.

It should be noted that F_α was not included in this final comparison as we thought it would complicate matters unnecessary to have the medical expert choose between three methods of which two were our own. We therefore did the comparison with the variation of our method which performed the best on the manual annotations (Section 2.4.1).

Method	$S_{k,\tau}$	F_α	F_σ
p_0	125	30	48
p_1	194	213	23
γ_0	-0.36	-0.34	-0.41
γ_1	-0.43	-0.31	-0.57
q	5.2	6.2	6.8

Table 2.1: Optimal parameters obtained using the training set and phantom. In addition k, τ, α, and σ were estimated to be 103, 4.6, 0.03, and 0.50.

2.3.6 Reproducibility and correlation with lung function

Airway centrelines and branch generations were extracted from the airway tree with a front propagation method, as described in Lo et al. 2012. Airway morphology was quantified using Lumen Diameter (LD) and Wall Area percentage (WA%) computed from distances of the inner and outer surfaces to the nearest point on the centreline. Measurements were averaged by branch generations, with trachea assigned generation zero. We assume that changes in the airways due to disease are relatively minor in the roughly one year period between the repeated scans, and that changes in the measures are mostly due to measurement variability. We quantify the reproducibility of the measures, with the coefficient of determination, calculated from Pearson product moment correlation coefficients.

2.4 Results

Table 2.1 shows optimal parameters obtained using the training set and phantom as described in Section 2.3.2 and 2.3.4.

Figure 2.3 illustrates the effect of the different column construction approaches on what constitutes "smooth" solutions in terms of the implemented penalties. It also illustrates the issues $S_{k,\tau}$ has with shorter columns in high curvature areas.

Running the methods on an image from our data usually takes less than 5 minutes (On a 1.6 GHz laptop using no parallelisation) using up to 4 GB of memory. Figure 2.4 shows a visualization of a three-dimensional segmentation result of F_σ and Figure 2.5 shows cross-sections illustrating results of the investigated methods.

2.4.1 Comparison with manual annotations

Table 2.2 shows the results of the comparisons with the manual annotations in the test data set, for each of the investigated methods using the optimal parameters. All methods performed well; no method had an average relative

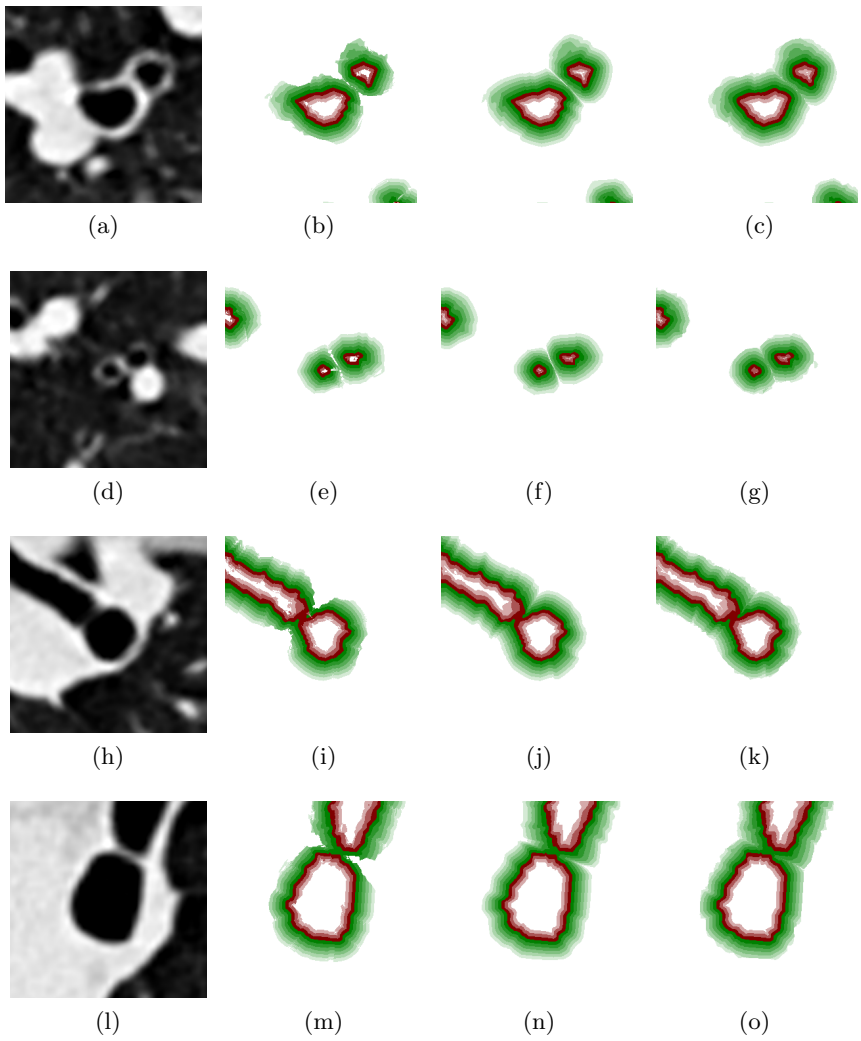
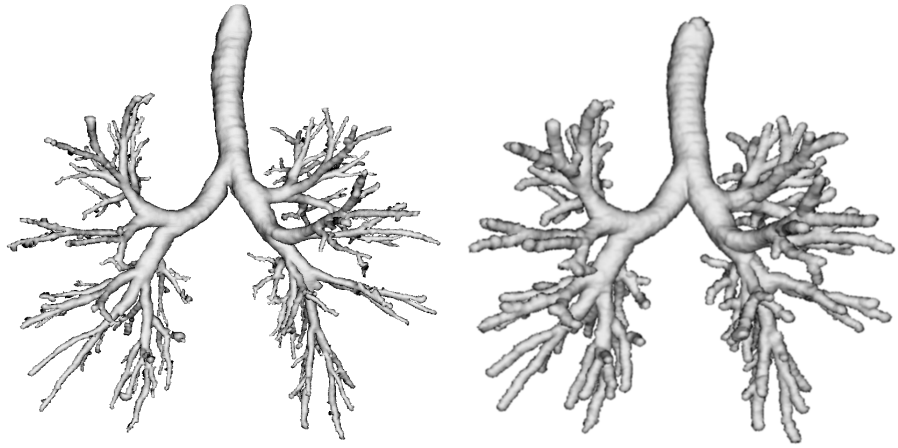


Figure 2.3: Identically coloured pixels illustrate solutions with zero smoothness penalty at different column indices. The outer-most red surface is the initial segmentation surface. The first column of images is the original scan, the next are of the $S_{k,\tau}$, F_α, and F_σ methods respectively. $S_{k,\tau}$ is most notably different from the flow line approaches in the high curvature areas, where a decreased smoothness can be observed, likely due to shorter columns. The surfaces of F_σ are different from F_α due to an increased local regularisation with limited long range effects, which is especially apparent in the bifurcation regions, where the surfaces of the two branches join earlier.



(a) Interior surface (b) Exterior surface

Figure 2.4: Three-dimensional visualizations of an inner and outer surface extracted by F_σ.

Method	Φ	Ψ (mm)
$S_{k,\tau}$	0.871 ± 0.075	0.135 ± 0.149
F_α	0.883 ± 0.066	0.110 ± 0.134
F_σ	$\mathbf{0.893 \pm 0.058}$	$\mathbf{0.096 \pm 0.097}$

Table 2.2: The results of different methods and kernels on the test data set with 319 manually annotated slices. Mean ± standard deviation of (2.19) and (2.20). The best result marked with a bold font.

area of overlap of less than 0.871 and a maximum average curve distance of more than 0.135 mm. F_σ achieved the significantly best result ($p < 0.05$) and the smallest variance in the quality of the results ($p < 0.05$) in terms of both metrics, while results of $S_{k,\tau}$ were the worst ($p < 0.001$). Results were compared using a Wilcoxon signed-rank test and a two-sample F-test respectively. It should be noted that these tests assume the samples to be independent, which can be questioned given that multiple cross-sectional slices were selected from each of the scans. However the mean accuracy of F_σ was better than $S_{k,\tau}$ in every single subject and using both metrics, as seen in Figure 2.6, which is significant in itself ($p < 0.05$).

2.4.2 Comparison with phantom

Table 2.3 shows the results of the phantom segmentation. Interior and exterior radii were estimated to within an average unsigned error of 6.6% (-15.4 % to 7.6 %) and 4.0 % (-8.0 % to 1.5 %) respectively. In absolute terms the radii

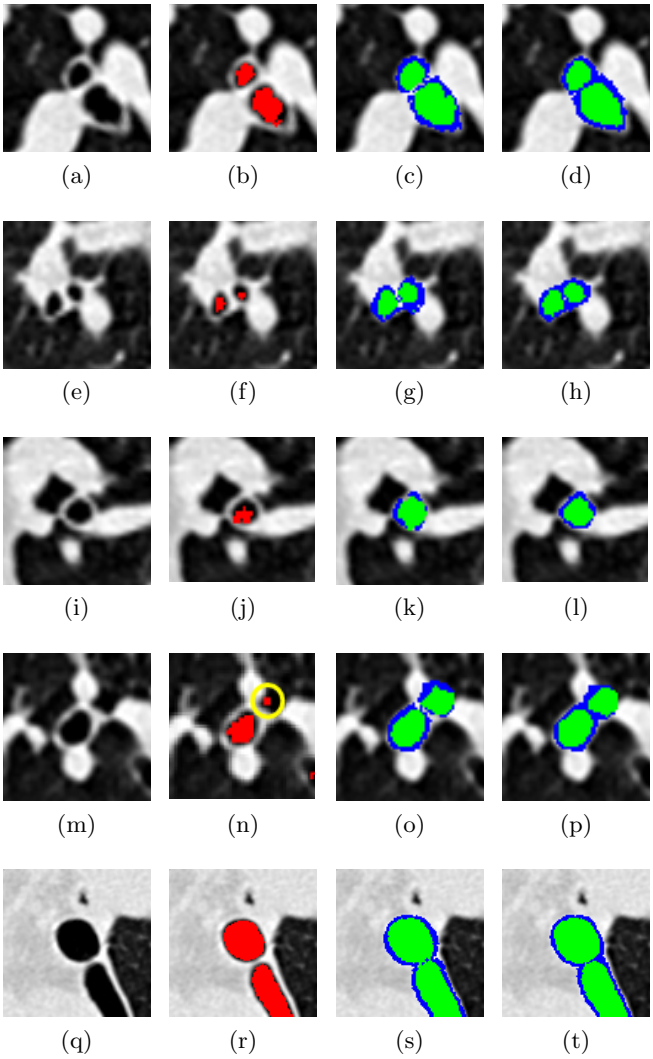


Figure 2.5: Cross-sectional images extracted perpendicular to airway centreline. Left to right show original image and initial, $S_{k,\tau}$ and F_σ segmentations. Only lumen (green) and wall (blue) segmentations connected to the centred airway are shown. F_σ is seen to do better than $S_{k,\tau}$ in the high curvature area between the bifurcating airway branches (first and second rows). The less noisy and more smooth result of F_σ compared to $S_{k,\tau}$ (third row) shows the flow line approach dealing better with the poor initial segmentation. The second to last row illustrates how neither method is able to correct large inital segmentation errors, such as the encircled over-segmented area. The last row shows a result near the mediastinum, where the outer border has weak contrast. In such situations smoothness penalties help by integrating information from nearby areas where contrast is stronger.

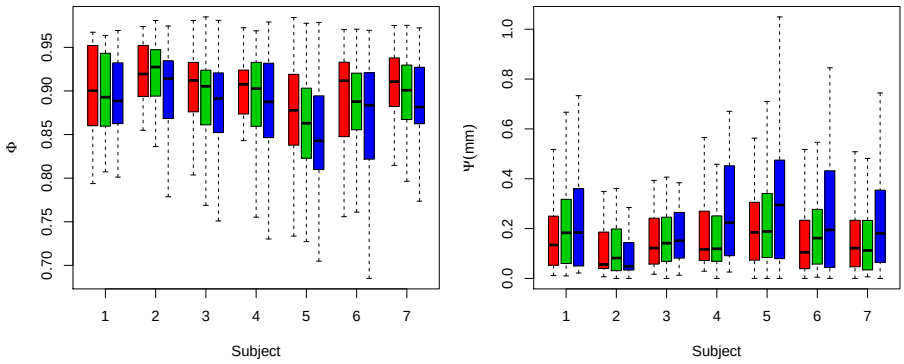


Figure 2.6: Result on the manually annotated test data set by subject of F_σ (red), F_α (green) and $S_{k,\tau}$ (blue). Bar shows median, box at lower and upper quartiles, and whiskers at interquartile range.

	Interior radius (mm)		Exterior radius (mm)	
Tube	Measured	Actual	Measured	Actual
0	3.23	3	4.14	4.5
1	3.00	3	3.80	3.9
2	1.27	1.5	2.13	2.1
3	1.33	1.5	2.11	2.1
4	3.10	3	3.98	4.2
5	3.07	3	3.94	4.2

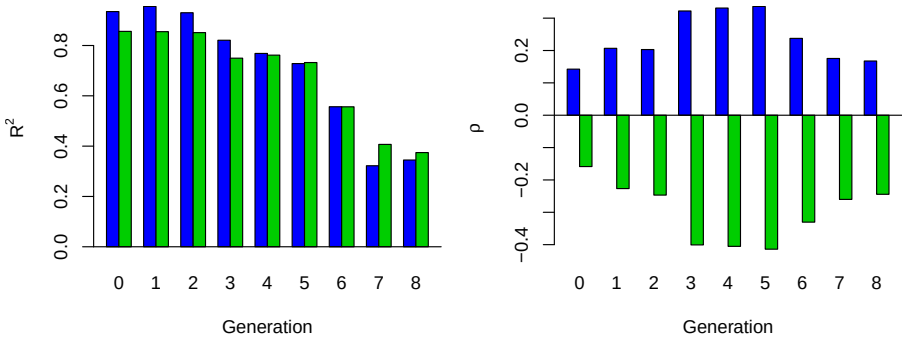
Table 2.3: Measured and actual phantom radii in mm.

deviated an unsigned average of 0.13 mm (-0.23 mm to 0.23 mm) and 0.16 mm (-0.36 mm to 0.16 mm) respectively, which is well below the graph resolution used.

2.4.3 Medical expert observer comparison

Out of the 499 evaluated cases the medical expert judged the flow line approach to be best in 342 cases, which is significantly more than the 56 cases in which the approach of Liu et al. 2012 was judged to be best, with $p < 0.001$ according to a proportion test. In 92 cases results were found to be of equal quality and in 9 cases the expert ruled that not enough information was present to judge.

The distribution of the proportion of cases where the flow line approach was judged to be best with respect to relative position of the cross-section within the branch was investigated to reveal whether the beginning, top 20% of the centreline; ending, bottom 20%; and middle, the rest of the branch was handled better than with the approach of Liu et al. 2012. The observer



(a) Reproducibility (b) Lung function

Figure 2.7: Figure 2.7a shows the reproducibility of the measures, LD (*blue*) and WA% (*green*), in repeated scans quantified as R^2 in branch generations 0 to 8. Figure 2.7b shows significant ($p < 0.01$) Spearman correlation coefficients of the same measures with lung function.

preferred the proposed approach in significantly more cases in all parts of the branch according to a proportion test ($p < 0.001$), specifically, 71 out of 107 cases (66%) in the beginning, 211 out of 301 cases in the middle (70%) and 60 cases out of 82 in the bottom (73%). The proportions were not significantly related to branch position ($p = 0.59$).

2.4.4 Airway abnormality measurements

Figure 2.7a shows reproducibility of LD and WA% quantified as repeat scan R^2. In general both measures have good reproducibility down to generation 5 ($R^2 > 0.73$) and the reproducibility is falling with increasing generation.

Figure 2.7b shows Spearman's correlation coefficients ρ of the measures and lung function measured by Forced Expiratory Volume in one second as a percentage of the predicted value (FEV1 (% predicted)). LD is positively correlated, indicating luminal narrowing with COPD, whereas WA% is negatively correlated, indicating wall thickening. Correlation is seen to increase with generation until generation 5.

2.5 Discussion

In this chapter we have shown how optimal net surface graphs based on columns defined from properly generated flow lines can be used to accurately find multiple interacting surfaces in multiple dimensions. Compared to previously used straight columns, the method should be better able to handle high curvature regions and noisy initial segmentations. We have demonstrated how the approach can be used to find the surfaces of the airway walls in CT images.

The high reproducibility of the resulting airway abnormality measurements as well as the fact that the measurements show significant correlation between decreased lung function and luminal narrowing and wall thickening, in agreement with the current knowledge of the disease process in COPD, indicate the method can be used to measure abnormalities caused by COPD up to at least generation 5.

We experimented with two different convolution kernels to obtain the flow lines, both giving good results. For our specific application, the Gaussian kernel F_σ performed better than F_α, which we think is mainly due to it tending to zero much faster, limiting long range effects while providing strong regularization near the potentially noisy initial surface. The most suitable kernel and kernel parameters may vary per segmentation task. In cases where one of the true surfaces can be far from the initial segmentation for instance, a kernel with longer tails may be needed.

In this work, we performed an extensive parameter tuning for each of the methods to allow for an as objective comparison as possible. However, it is our experience that results are not very sensitive to the settings of these parameters and suitable settings can already be obtained by a few trial and error runs using visual inspection of the segmentation results.

The graph resolution was set relatively low (0.5 mm), compared to that of Liu et al. 2012 (0.1 mm) to limit computation times and memory consumption. This results in a slightly lower accuracy on phantom scans than what Liu et al. 2012 report, however, there is no reason to doubt if the improvement in accuracy on real data as seen with our approach will carry over to higher graph resolutions as well.

The chosen mesh is not very smooth and so it could be questioned whether the methods perform differently simply because of differences in their ability to deal with this non-smoothness. To explore whether this was the case we repeated the training (Section 2.3.2) with meshes based on the marching cubes algorithm, as was used in Liu et al. 2012 and a smooth mesh based on Boissonnat and Oudot 2005 and evaluated their results on the test set (Section 2.3.3). The mesh based on Boissonnat and Oudot 2005 overall performed similar to our mesh, whereas marching cubes performed significantly worse with each of the investigated methods. Moreover the ordering of the methods in terms of results seemed to be preserved. This indicates that our mesh choice is not the cause of the differences. An advantage of our mesh is that each vertex has a constant number of adjacent vertices (four) and the distance to its neighbours varies less than with the tested alternatives. This regularity helps make the graph smoothness penalties work more evenly across the surface.

Estimating both airway wall surfaces simultaneously using the initial lumen surface as a prior, makes sense, as the lumen surface is easier to segment than the outer wall surface. This is because the contrast between the outer wall surface and abutting structures, such as vessels can be very low. In low contrast areas results become more driven by smoothness penalties. If the

initial segmentation is not unreasonable, good segmentations can be achieved even in these regions as seen in Figure 2.5d, 2.5h and 2.5t, however in cases where contrast is low and the initial segmentation is also poor, errors can occur. Examples of this can be seen in Figure 2.5l, where the wall area in the vascular contact region seems to be underestimated. Sensitivity to the accuracy of the initial segmentation is however not unique to the proposed approach, as all previously developed three-dimensional airway wall segmentation approaches (Liu et al. 2012; Saragaglia et al. 2006) depend on an initial segmentation and one and two-dimensional methods require an accurate estimation of the centreline (Petersen et al. 2010; Weinheimer et al. 2008; Estépar et al. 2006). The method we used to obtain the initial segmentation is state-of-the-art and has been shown to have very few false positives (Lo et al. 2009). Moreover as the lumen segmentation in general seem to be improved compared to the initial segmentation, it is possible that further improvements could be achieved by multiple iterations of the proposed approach.

Although manual annotations can be biased (King et al. 2000), comparison with them is still a valid way to assess a method's ability to find the airway surfaces in a realistic setting. Manual tracings were used in a similar fashion in Li et al. 2006 in 39 randomly selected slices outside the bifurcation areas. The unsigned errors were reported to be 0.10 ± 0.11 and 0.12 ± 0.12 mm for the inner and outer surfaces respectively, which is similar to what our approach achieved.

An ability to pick up significant changes in airway dimensions related to disease has, to our knowledge, not been demonstrated by previous fully automatic three-dimensional airway wall segmentation approaches. However, other one and two-dimensional approaches have, and significant correlation have for instance been reported with lung function using the approach of Weinheimer et al. 2008 in Achenbach et al. 2008, using the full width at half maximum principle in Nakano et al. 2000 or with exercise capacity in COPD using the approach of Estépar et al. 2006 in Diaz et al. 2010b. One should exercise care in drawing conclusions from comparisons of these studies as there are important differences in patient characteristics, and the number of healthy participants, however, our results are within the range reported by them (WA% coefficients of -0.338 (Nakano et al. 2000) to -0.560 (Achenbach et al. 2008)).

Scan-rescan repeatability of the measures was good, but did decrease from generation 6 and onwards as can be seen in Figure 2.7a. A large part of this decline can probably be explained by missing branches in the initial segmentations, leading to different branches contributing to the airway measures per generation at the different time points. Comparing the number of segmented branches to the expected value, assuming a binary tree structure, revealed that almost all the branches were segmented in generation 5 (97%), whereas the number had dropped to about 62% in generation 6. Measurements conducted in corresponding branches, as opposed to generations, might thus still be reproducible after generation 5, as was also seen in Petersen et al. 2011a

where image registration was used to do branch matching and limit measurements to repeatedly found branches. The use of higher dosage and resolution scans would probably also result in a higher number of detected branches and higher accuracy in placing the airway wall borders, due to a better definition of edges in especially the smaller airways. This should further improve reproducibility and correlation with lung function especially beyond generation 5.

We expected to see a higher frequency of cases where the proposed method was found superior in the high curvature areas near the bifurcations. The medical expert visual inspection did show a slightly higher proportion in the ending compared to the top and middle parts of the branches, but the difference was not significant. The results show the proposed method actually improves many cases outside bifurcations as well. An example is given in Figure 2.5. This could be due to small errors in the initial segmentation, which make it difficult to obtain good estimates of the normal directions. In such cases increased regularisation may help the flow line approach and even if the flow lines follow the same erroneous directions close to the initial segmentation, as they move away, the directions will be relatively more determined by long range interactions, and thus the errors will be evened out.

2.6 Conclusions

To conclude, a new graph construction technique applicable to multi-dimensional multi-surface segmentation problems was proposed. The method runs in polynomial time and is able to penalize for non-smoothness and separation of the found surfaces. The results are guaranteed to not self-intersect and are robust in regions with large curvature.

We applied the method to the problem of segmenting human airway walls in CT images based on an initial coarse airway lumen segmentation, and results were shown to be significantly more accurate than those of another recently published graph based method. Phantom dimensions were estimated to sub-voxel resolution, and large scale evaluations on 980 images from a lung cancer screening trial showed both good reproducibility of the obtained airway abnormality measures and a significant correlation with lung function.

Acknowledgements

This work was partly funded by the Netherlands Organisation for Scientific Research (NWO), and AstraZeneca Sweden.

Chapter 3

Branch matching for longitudinal studies

The work presented in this chapter is based on J. Petersen, V. Gorbunova, M. Nielsen, A. Dirksen, P. Lo, and M. de Bruijne (2011a). "Longitudinal Analysis of Airways using Registration". In: *Fourth International Workshop on Pulmonary Image Analysis*. Ed. by R. Beichel, M. de Bruijne, B. van Ginneken, S. Kabus, A. Kiraly, J. Kuhnigk, J. McClelland, K. Mori, E. van Rikxoort, and S. Rit. CreateSpace, pp. 11–22.

Abstract

Longitudinal investigations of airway abnormalities associated with Chronic Obstructive Pulmonary Disease (COPD) has been very limited so far, partly due to the difficulties in obtaining reproducible measures.

We propose to improve on this by limiting measurements to corresponding branches found using image registration.

The results obtained from scans of 237 subjects show increased intra-subject correlation when measurements are conducted in branches found in each scan compared to similar measurements not limited to corresponding branches. This indicates the method could be useful for longitudinal analysis.

Yearly changes in CT measures showed that airways increase in size and decrease in density with time. Changes were in general not found to be significantly correlated with changes in lung function and neither were there any significant differences between COPD GOLD stages.

3.1 Introduction

Chronic Obstructive Pulmonary Disease (COPD) is associated with loss of lung tissue, known as emphysema and chronic bronchitis, which is normally described as a narrowing of the air-filled lumen area of the airways and a thickening of the airway walls (Hogg et al. 2004). The changes cause shortness of breath leading to reduced quality of life, disability and eventually death. Computed Tomography (CT) has become a popular imaging tool to quantify COPD pathology and multiple cross-sectional studies have already shown that CT based measures of both emphysema and airway abnormality are correlated with measures of disease severity such as lung function (Berger et al. 2005; Hasegawa et al. 2006; Nakano et al. 2000; Washko et al. 2009). As of yet only very few longitudinal studies have attempted to investigate the change over time in airway dimensions (Ohara et al. 2008). Some of the reasons for the lack of longitudinal studies are likely that COPD is a very slow developing disease, meaning subjects need to be followed for many years and the lack of reliable automatic approaches for measuring airway dimensions.

Conducting reproducible measurements in the airways is difficult as such measurements in general depend on the position in which they are performed and finding the same position in the following images can be a hard task. For instance in the longitudinal study described in (Ohara et al. 2008), such correspondences were found manually by locating the anterior, lateral and posterior basal segment bronchus in CT slices. In the end data on 45 out of the 83 subjects participating had to be excluded as the same segment bronchus could not be located in at least three of the yearly scans. The use of modern tools such as three-dimensional segmentation algorithms, centreline extraction and axial reconstruction enable analysis in much more of the airway tree, making the problem of how to do comparable measurements even more complicated.

One way to avoid this problem is to use airway abnormality measures, which are less affected by differences in sampling positions. For instance one can use the assumption that the square root of the wall area and the lumen perimeter are linearly related, when measured in perpendicular slices of the airways, to construct a comparable measure of what the wall area would be if measured where the lumen perimeter is 10 mm. This is the much used Pi10 measure (Nakano et al. 2005). Such approaches probably increase reproducibility, however measurements are likely still dependent on the specific branches included in the analysis.

The airway branches have been given anatomical names down to the sub-segmental level within the literature. If such a labelling could be extracted automatically it would enable measurements in correspondingly labelled branches, that would be comparable cross-sectionally as well as longitudinally. Assigning these names is however very difficult in practice, due to biological variation, inspiration effects, pathology, etcetera. So the automatic processes that have been developed usually only proceed down to the 10 seg-

mental bronchi on each side, resulting in 32 labelled branches (Ginneken et al. 2008b; Tschirren et al. 2005b).

Modern segmentation methods on data of reasonable image quality can go significantly deeper than the segmental bronchi and it thus might be possible to match more branches at an intra-subject level. This could be useful in longitudinal studies where the inter-subject variation often is less relevant and particularly important for COPD analysis as it is known to affect the airways further down the tree more (Hasegawa et al. 2006). One way to achieve such a matching would be to use image registration. Registration of lung CT images has, for example, been used to track emphysema progression (Gorbunova et al. 2008) and nodule growth (Zheng et al. 2007), however so far, to the best of our knowledge, it has not been used to investigate airway changes.

The purpose of the work detailed in this chapter is thus to investigate longitudinal measurements of airway abnormalities and whether limiting measurements to branches only found in each intra-subject scan, matched with the help of image registration, improves reproducibility.

3.2 Method

A fully automatic and novel framework for longitudinal analysis of changes in airway wall dimensions and density was developed. It uses state-of-the-art airway and airway wall segmentation and registration methods, described shortly in the following sections. Briefly: the airway lumen was initially segmented using the process described in Sec. 3.2.1, it was then used as input to the airway wall segmentation method detailed in Sec. 3.2.2 in order to find the precise shape and position of the airway wall surfaces. The lumen surface returned from this was used to find the airway centrelines using the process described in Sec. 3.2.3. The airway centrelines were deformed to the centre-most image in time using the deformation fields returned by the image registration process described in Sec. 3.2.4. This common space allowed the centrelines to be matched based on distance and orientation using the method detailed in Sec. 3.2.5.

3.2.1 Initial airway extraction

The airway segmentation method described in Lo et al. 2009 that iteratively extends locally optimally paths to form an airway tree is used in this work. In each iteration, locally optimal paths are defined as paths with minimal cost from the seed-point to the surface of a sphere centred on it. The paths are generated using Dijkstra's algorithm, with a cost function that is based on a kNN classifier trained to classify airway voxels combined with Hessian eigenanalysis to enhance cylindrical structures. A number of criteria taking into account the local appearance of an airway voxel and geometry characteristics are then used to select the most likely path. The paths are then converted into

a full lumen segmentation by growing a cylinder around each selected path, using the airway probabilities returned by the kNN classifier.

This airway segmentation method was chosen as it compared very favourably with another region growing based approach (Lo et al. 2008), which again performed well compared to the state of the art evaluated in the Exact'09 study (Lo et al. 2012). One of the advantages of the approach is that it can overcome local occlusions, due to for instance plugging of the lumen due to mucus or pathology.

3.2.2 Airway wall segmentation

The initial lumen segmentation is then used as input to the method described in Chapter 2, which builds an optimal surface graph (Wu and Chen 2002) around it with the purpose of both finding the outer airway wall surface and refining the lumen surface returned by the first step. This process begins by converting the initial lumen segmentation into a sub-graph, in which each point on the initial lumen surface is associated with a column of nodes. A column defines the set of allowed positions the point can take in the sought surface. Optimal surface graphs are designed such that the search for the optimal surface can be conducted using maximum-flow/minimum-cut algorithms in polynomial time. The process can be thought of as a refinement or a deformation in which the sub-graph defines the finite set of possible refined solutions of the initial surface. Since we need to find both the inner and outer airway wall surface the complete graph consists of two sub-graphs, one designed to find the inner surface and one designed to find the outer. The optimality of the solution is measured in terms of inner and outer surface cost functions computed from derivatives of the image intensities as described in Petersen et al. 2010 and smoothness and surface separation priors.

One of the novelties of the algorithm is the way the graph columns are constructed from properly generated greatest ascent and descent flow lines. These guarantee solutions that do not self-intersect and should be very suited for regions with high curvature, such as those found in the branch bifurcation areas.

3.2.3 Extraction of branch centrelines and generations

The airway centrelines, branches and generations were extracted from the lumen surface generated in the airway wall segmentation process using the front propagation method described in Lo et al. 2012. Starting in the trachea and moving down the branches, the centroid of the front is stored at regular intervals as branch centreline points. Bifurcations are detected and generation count increased as the wavefront becomes disconnected upon hitting the branching points of the lumen segmentation. The method was also used in

the Exact'09 study and had thus already been used on a varied data set and on the results of different airway segmentation algorithms.

3.2.4 Registration

The extracted centrelines were matched within a common coordinate system obtained using registration of the CT images.

Image registration is the process of finding a transformation which maps one image into another. This is usually performed in a pairwise manner, where one image is denoted the moving image and the other the fixed. The transformation maps the coordinates of the fixed image into the moving.

The registration error is generally related to how dissimilar the images are and since the data set consists of subjects scanned five times yearly, we assumed the centre-most image in time to be least different from the others and registered this image as the moving image with all the four others.

The images were registered using the mass preserving image registration algorithm described in Gorbunova et al. 2008. Registration based directly on image intensities, such as standard sum of squared differences, is problematic for lung registration because the local image intensity, which in CT images are directly related to the local density, changes with the inspiration cycle. Instead the approach incorporates a tissue appearance model based on the assumption of preservation of total lung mass into a standard deformable image registration framework. This framework uses a composition of a global affine and three free-form B-Spline transformations with increasing grid resolution. A version of sum of squared differences with the mass preservation incorporated is used as a similarity function.

The method was originally evaluated using the average distance between the registered lung vessel trees, and showed a significant improvement compared to standard sum of squared distances, especially in the more difficult cases with large differences in lung volume.

3.2.5 Matching airway branches

Each centreline point was matched to the nearest point on each of the other centrelines of the same subject, measured within the common coordinate system. Such a match was deemed acceptable if the distances to the common centre for each of the centreline points was less than δ and the angle the direction of the centrelines formed with the average direction of the centrelines, was less than θ. The airways were then cropped at the position where the deepest acceptable match was found.

Having accurately assigned generation numbers is important as COPD mostly affects the smaller airways, and splitting measurements by generations is a common pathology independent way to only measure relevant airway branches. See for instance the results of the generation based analysis con-

ducted in Hasegawa et al. 2006. Most erroneously assigned generations are due to segmentation errors, where only one of the continuing branches at a bifurcation is found. In these cases one longer branch may be found where it should have bifurcated and split in two and the sub-tree will have its generation count off by one. The matched centrelines were used to correct some of these cases. Beginning at the root of the tree and moving down the centreline the current generation is determined by a majority vote. Missing or spurious branches are detected whenever one of the centreline points have a different generation number than the majority. It is corrected by adding the difference to each of the centreline points in the sub-tree. The process allows a correct identification of the generations, even in situations where errors occur in all the trees from any string of branches from root to leaf, as long as each bifurcation is found in the majority of the cases. See Fig. 3.1c for a visualization of the results of this.

3.3 Experiments and results

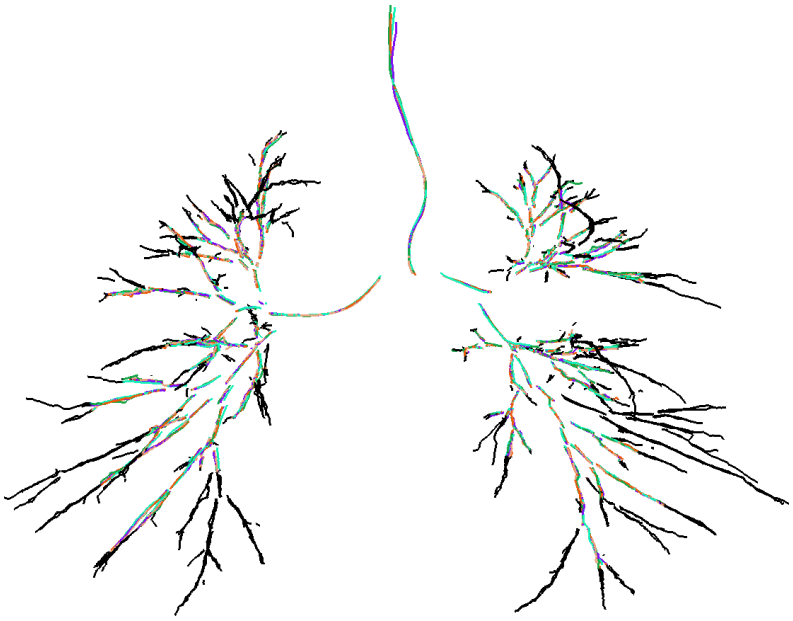
3.3.1 Data

The material used comes from the Danish lung cancer screening trial (Pedersen et al. 2009). The images were obtained using a Multi Detector CT scanner (16 rows Philips Mx 8000) with a low dose (120 kV and 40 mAs), reconstructed using a hard kernel (D) with a resolution of approximately $0.78mm \times 0.78mm \times 1mm$. 237 randomly selected subjects from the trial were included in the analysis. Data from five yearly CT scans and lung function measurements were available on each.

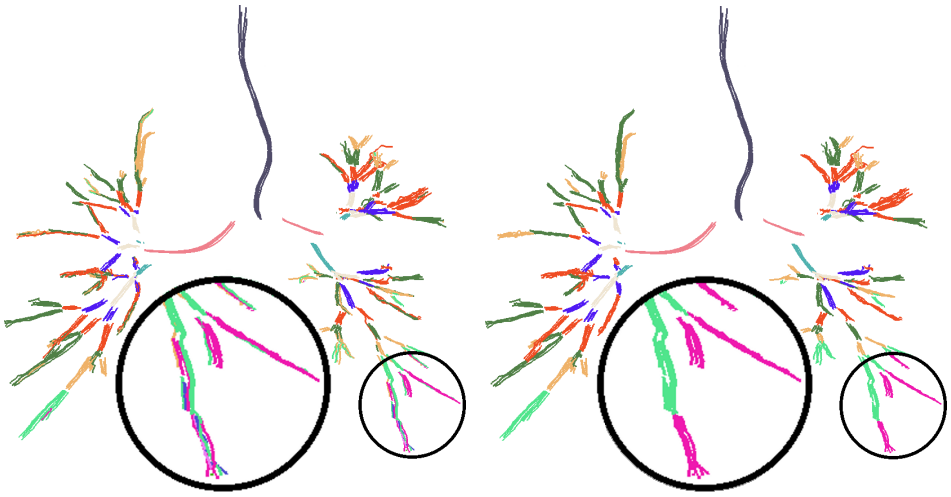
At baseline, the subjects included in the analysis had a mean value of FEV1 (% predicted) of 96%($\pm$17%) and FEV1/FVC of 0.71($\pm$0.08), totalling 144 without COPD, 61 with mild COPD, 31 with moderate COPD and 1 with severe COPD. 143 men and 94 women with an average age of 58($\pm$5) years.

3.3.2 Centreline cropping

A value of $1.7mm$ for δ and 37 degrees for θ were estimated by visual inspection on a small set of scans that are independent of the data used in the rest of the analysis. Fig. 3.1 shows a typical result of running the branch cropping method. Notice, in Fig. 3.1a how the five deformed centrelines are so close that they more or less appear as one single airway centreline, indicating how well the registration method works on this data. Fig. 3.1b and 3.1c shows the results of the generation correction method. The improvement is most clearly visible in the bottom of the tree, where the many missing branches leads to wrongly detected generations, most often seen as the same branch being coloured with multiple colours in Fig. 3.1b.



(a) Five centrelines



(b) Individual Generations (c) Majority generations

Figure 3.1: Fig. 3.1a shows the centrelines deformed to a common space, each tree has a unique colour, the cropped parts of the tree where at least one scan is missing a branch are black. Fig. 3.1b shows the trees coloured by the generations of the individual trees and Fig. 3.1c shows the corresponding majority generations. The enlarged area within the circles shows branches where generations are corrected. Note that in the last two figures each tree has been slightly offset compared to the others, such that the generations of each individual tree is more easily visible.

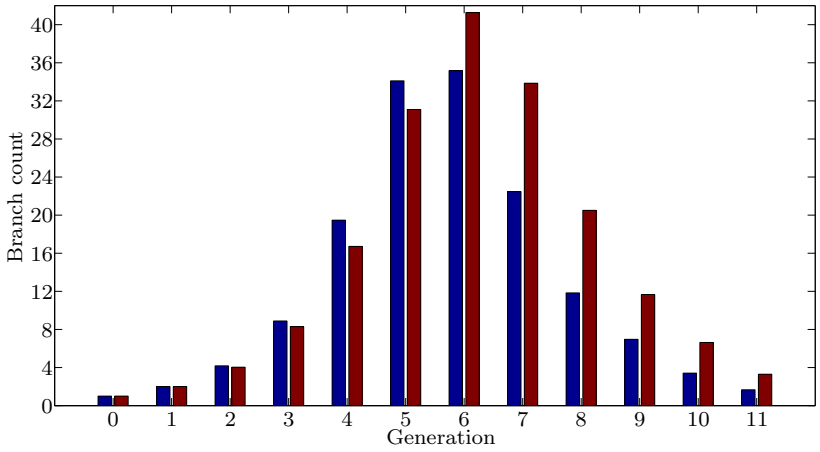


Figure 3.2: The average amount of branches in each generation of the cropped (blue) and non-cropped (red) airways.

Fig. 3.2 shows the amount of branches found in each generation of the airway tree in both the cropped and non-cropped case. The amount of branches in the cropped airways are actually larger in generation 3-5 compared to the non-cropped airways due to the generation correction process. Moreover the number of branches in each generation are roughly doubled in both cases until generation 6, consistent with a bifurcating branching tree. This suggests that the airway trees are roughly complete until this point and that the cropped branches mostly belong to the generations that are already incompletely found. It is also interesting that even with as many as five time points the average number of branches found in the cropped airways far exceeds the number of named branches.

3.3.3 Measurements

Airway morphology was quantified with four different measures, the Interior Volume (IV) (also called the lumen volume), the complete Airway Volume (AV), that is the sum of the interior and wall volume, the wall volume percentage (WV% = 100 × (AV-IV)/AV) and the Mean Airway Density (MAD), which is the average density in the complete airway volume. Density based measures have received some attention in the last couple of years as they may be more sensitive to a change in size of the smaller airways due to partial volume effects (Petersen et al. 2010; Washko et al. 2009). IV, AV and WV% on the other hand are three-dimensional extensions of commonly used airway abnormality measures (Berger et al. 2005; Hasegawa et al. 2006; Nakano et al. 2000; Ohara et al. 2008; Petersen et al. 2010; Washko et al. 2009).

The measurements were conducted in the un-registered images, by classi-

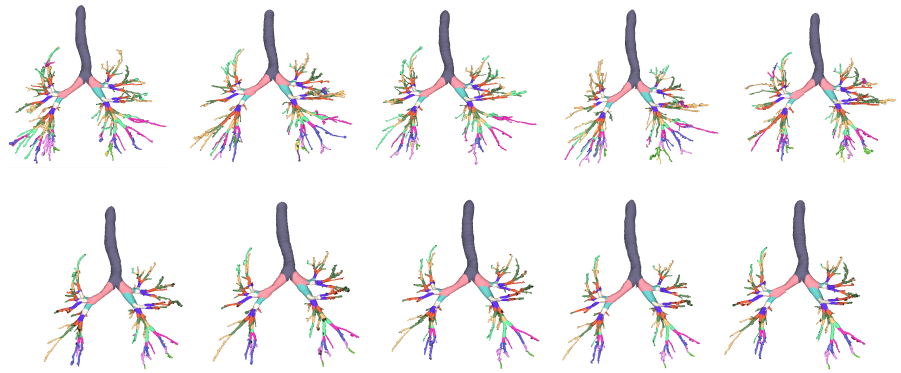


Figure 3.3: Lumen segmentation surface in five scans of one subject. Top row shows all the branches, whereas the bottom row only show corresponding branches.

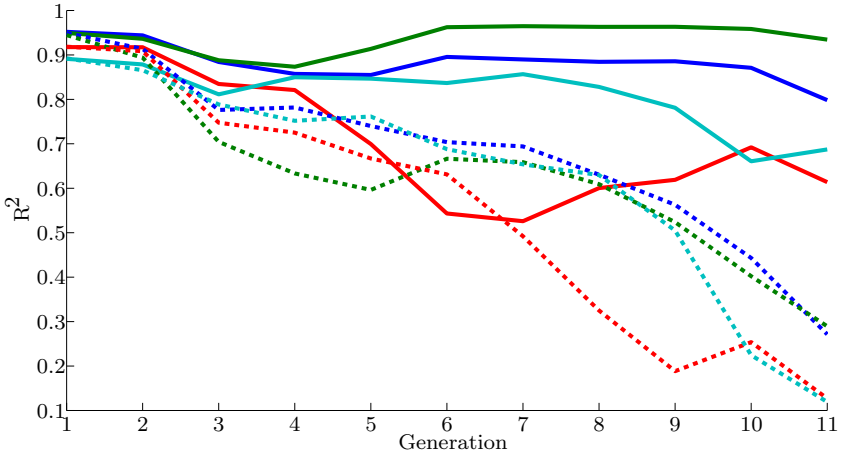


Figure 3.4: Reproducibility as the coefficient of determination calculated from measurements at baseline and first year repeat scan. Red, green, dark and light blue is WV%, AV, IV and MAD respectively. The dotted lines are measurements based on the non-cropped airway trees.

fying the segmentation into branch generations with the use of the cropped centrelines. This was done by assigning the generation of the nearest centreline point to each segmented voxel as described in Chapter 2. Voxels whose nearest centreline point was cropped, were simply not included in the measurements. Fig. 3.3, shows an example of this.

	IV (mm^3)	AV (mm^3)	WV% (%)	MAD (HU)
Mean (± std)	55 (±547)	328 (±1120)(***)	0.026 (±0.51)	-2.9 (±7.4) (***)
FEV1 (%)	-0.04	-0.08	-0.00	-0.12
COPD FEV1 (%)	-0.01	-0.07	-0.03	-0.21(*)

Table 3.1: Change in measurements per year. Second and third row are Pearson correlation coefficients. Third row are measurements conducted only within subjects diagnosed with COPD at baseline. Note /year has been left out from the units for readability. (*), (***) denote p values less than 0.05 and 0.001 respectively.

3.3.4 Reproducibility of measurements

COPD is a slow developing disease and so airway abnormality measurements can be assumed to change little from one year to the next. The coefficient of determination calculated from the Pearson correlation coefficient of the measures at baseline with the following year can thus be used to estimate reproducibility of the measures. Fig. 3.4 shows how this looks in each generation for the measures extracted from the cropped and non-cropped airways.

It is clear that the cropping operation results in more reproducible measures, especially in the smaller branches, with the only exception being WV% measured in generation 6. In general the reproducibility in the non-cropped airways can be observed to fall with generations, the fact that this trend has largely been removed after the cropping operation suggests that a large part of this variability can be attributed to differences in the amount of segmented airway branches.

3.3.5 Annual change in measures

The first row of Table 3.1 shows the annual change in the measures obtained via the slope of the linear relationship between subject age at the time of the scan and the measurements in the airway belonging to generation 3 and up. IV and WV% showed no significant annual change, whereas AV was found to increase and MAD to decrease, indicating that the airways increase in size and become less dense.

The annual change of the measures was not found to be correlated with the annual change of lung function. However when analysis was limited to the subjects with COPD at baseline, a significant negative correlation could be observed between the annual change in MAD and the annual change in FEV1(% predicted). It should be mentioned though that the result is no longer significant if the level is Bonferroni corrected. The result is however consistent with cross-sectional studies, which found that poorer lung function

in general was associated with a higher density (Petersen et al. 2010; Washko et al. 2009).

We also tested whether there were any significant differences between the different GOLD stages of COPD severity for any of the measures, but found none.

3.4 Discussion

The implemented branch matching procedure is simple compared to anatomical labelling approaches (Ginneken et al. 2008b; Tschirren et al. 2005b), but visual inspections indicate that it works well, likely because the registrations are good. That is, the distance between corresponding branches and their mutual angles within the common coordinate system are generally smaller than the distances and angles to non-corresponding branches.

Compared to recent cross-sectional studies (Berger et al. 2005; Hasegawa et al. 2006; Nakano et al. 2000; Washko et al. 2009), where wall thickness and lumen area or volume have been found to be correlated with poor lung function, it is perhaps surprising that this relationship wasn't reflected in our longitudinal measurements. The lack of correlation can however be explained by the fact that COPD develops slowly, and thus five years might simply be too short to see any significant change in this data set. Moreover both lung function and CT based measurements are still very noisy and the registration based cropping operation described in this chapter, is only able to reduce the intra subject variation introduced by differences in the amount of found airway branches. Other variations, such as for instance those caused by changes in inspiration level are still influencing the measurements. It should be mentioned that in the results presented in Ohara et al. 2008, a significant correlation was found between changes in airway measurements and lung function on a smaller data set. However, that study included more severe COPD cases, with a baseline mean value FEV1/FVC of 0.51 and FEV1(% predicted) of 50%.

An advantage of using image registration over for example branch labelling methods, is that it allows for measuring changes of for instance wall thickness, density, etcetera, local to specific points in the airways. Such changes could be visualized to increase understanding of the disease or combined to form new, possibly more sensitive global measures.

3.5 Conclusion

A fully automatic framework for longitudinal analysis of airways was presented, using state-of-the-art airway wall segmentation and image registration methods.

The process of limiting measurements to airway branches found in each repeated scan was shown to increase reproducibility. This indicates that a large

part of the intra-subject variation in the measurements can be attributed to differences in the amount of segmented branches and thus that the framework could be useful for longitudinal studies. The number of matched branches exceeds the number of anatomically named branches.

A significant annual increase of AV and decrease of MAD was observed. Annual changes in the CT measures was not observed to be correlated with annual changes in FEV1(% predicted) in the complete data set, nor were there any significant differences between the means of subjects in the different COPD stages. However the change in MAD was seen to be negatively correlated with the change in FEV1(% predicted) when limited to the subjects with COPD at baseline.

Acknowledgements

This work is partly funded by the Netherlands Organisation for Scientific Research (NWO), and AstraZeneca, Sweden.

Chapter 4

Geodesic anatomical labelling

The work presented in this chapter is based on A. Feragen*, J. Petersen*, M. Owen, P. Lo, L. H. Thomsen, M. M. W. Wille, A. Dirksen, and M. de Bruijne (submitted). "Geodesic anatomical labeling of airway trees". In: -.

Abstract

We present a fast and robust supervised algorithm for labelling airway trees, using geodesic distances in a geometric tree-space. Possible branch label configurations for an unlabelled airway tree are evaluated using distances to a training set of labelled airway trees. In tree-space, airway tree topology and geometry change continuously, giving a natural automatic handling of anatomical differences and noise. A hierarchical approach makes the algorithm efficient, assigning labels from the trachea and downwards. Only the airway centreline tree is used, which is relatively unaffected by pathology. The algorithm is evaluated on 80 segmented airway trees from 40 subjects at two time points, labelled by 3 medical experts each, testing accuracy, reproducibility and robustness in patients with Chronic Obstructive Pulmonary Disease (COPD). The accuracy of the algorithm is statistically similar to that of the experts and not significantly correlated with COPD severity. The reproducibility of the algorithm is significantly better than that of the experts, and negatively correlated with COPD severity. Evaluation of the algorithm on a longitudinal set of 8724 trees from a lung cancer screening trial shows that the algorithm can be used in large scale studies with high reproducibility, and that the negative correlation of reproducibility with COPD severity can be explained by missing branches, for instance due to segmentation problems in COPD patients. We conclude that the algorithm is robust to COPD severity given equally complete airway trees, and comparable in performance to that of experts in pulmonary medicine, emphasizing the suitability of the labelling algorithm for clinical use.

4.1 Introduction

Computed Tomography (CT) is an important tool in the analysis of diseases affecting pulmonary airways. Using image segmentation methods, three-dimensional models of the airway surfaces can be constructed, and their dimensions measured. Measurements such as lumen diameter, airway wall thickness, and bifurcation angle are, however, dependent on the location in which they are made; e.g., in Hasegawa et al. 2006 it is shown that the classification boundary and accuracy of using airway wall thickness for COPD prediction is different in different locations in the lung. As a consequence, it is crucial to determine anatomically corresponding positions in different airway trees in order to robustly compare measurements across patients. One way to solve this problem is to identify the airway tree branches by their anatomical names, and compare measurements in identically named airway branches (Hasegawa et al. 2006; Lederlin et al. 2012; Diaz et al. 2010a). Identifying the anatomical names is non-trivial, since the topology of the airway tree changes from person to person, and the segmented trees have additional differences introduced by noise, including missing and spurious branches.

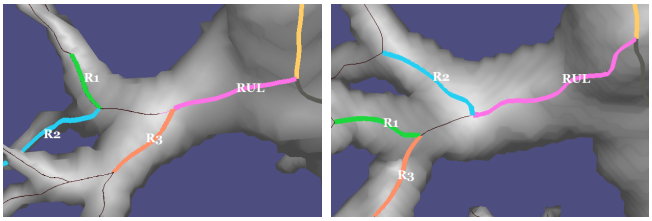


Figure 4.1: Airway trees are frequently topologically different, while geometric differences are small.

Several airway branch labelling algorithms have appeared in the literature. Gu et al. 2012 label the lobe branches by assuming a fixed tree topology and assigning labels based on the (x, y, z) coordinates of bifurcations, a method which is sensitive to topological differences and thus in particular does not extend to segment labels. Ginneken et al. 2008a, Mori et al. 2009 and Lo et al. 2011 assign branch labels using supervised learning on branch features, in some cases (Ginneken et al. 2008a; Lo et al. 2011) constrained by assumptions on airway tree topology. Among the features used are branch length, radius, orientation, cross-sectional shape and bifurcation angle. Branch radius is sensitive to diseases like asthma, cystic fibrosis, tuberculosis and Chronic Obstructive Pulmonary Disease (COPD) (Pu et al. 2012; Lederlin et al. 2012; Wielputz et al. 2013; Moon et al. 1997). Moreover, airway branch length, shape and bifurcation angle are sensitive to anatomical differences in topology and, in particular, missing branches in the airway segmentation. For instance, if only one branch in a bifurcation is detected, the result will be a longer branch with different shape and different endpoint bifurcation angles. It is known that fewer branches are typically detected in airway trees from subjects with COPD (Diaz et al. 2010a; Pu et al. 2012), leading to topological irregularities. These segmentation problems will affect any method which enforces constraints on airway tree topology.

Anatomical tree labelling is closely related with anatomical tree matching, or the problem of matching the branches or bifurcations of one tree to those of another, in the sense that matching an unlabelled tree to a labelled one will generate a label transfer to the unlabelled tree. Pisupati et al. 1996 use tree matching in airway trees of dogs. Graham and Higgins 2006 use a dynamical programming approach to graph matching for matching pairs of airway trees. Tschirren et al. 2005b and Kitaoka et al. 2002 label airway trees using association graphs for pairs of trees, which incorporate information from both trees, such that maximal cliques in the association graph induce branch matchings between the original graphs. A similar approach is used by Metzen et al. 2009 for matching both airway trees and vessel trees in the liver, as well as by Bogunovic et al. 2013 for labelling the Circle of Willis. While branch shape features go into the construction of the association graph, the possible

branch matches are subject to strict constraints as the matching is equivalent to identifying maximal isomorphic sub-trees. In particular, the association graph model is not able to take into account the way that arbitrarily small changes in geometric branch features (e.g. branch length) can lead to new topologies, as in Figure 4.1.

Other, more geometric approaches, also appear: Smeets et al. 2010 match branches from lung vessel trees using pairwise distances between bifurcations both in 3D Euclidean space and along the tree to generate distance matrix "fingerprints", which are matched. Bülow et al. 2006 match airway tree branches without connectivity information, using only branch shape context. Kaftan et al. 2006, match tree *paths* rather than branches, thus avoiding the difficulty with different tree-topological structures, but also losing all information stored in the topological structure. In particular, this model does not generate branch labels, as the branch division is lost.

Feragen et al. 2011 label airways based on geodesics, or shortest paths, in a space of trees. Their tree-space has no known efficient algorithm for computation of geodesics, making their method too computationally expensive to label the whole airway tree. In this work we use a more restrictive space of leaf-labelled trees (Billera et al. 2001), where geodesics can be computed in polynomial time (Owen and Provan 2011).

We present a novel supervised algorithm for automatic airway branch labelling, based on geodesic distances in a space of leaf-labelled trees. The labelling algorithm works by suggesting a set of branch label configurations, each forming a suggested labelled airway tree. These suggestions are compared using geodesic distances from each suggested labelled tree to airway trees within a training set labelled by clinical experts and the optimal suggested labelling is returned. Labels are thus assigned automatically from the trachea and downwards in a hierarchical fashion.

The tree-space framework is able to compute distances between trees with different topologies, allowing for flexibility in tree topology. This is possible because in tree-space, tree topology and branch geometry are allowed to change continuously. See Figure 4.2 for an illustration. From a practical point of view, this allows us to take advantage of the whole training set without being restricted by topological airway tree differences.

The only feature used by the labelling algorithm is the airway centreline tree, divided into branches. The algorithm does not depend directly on the division of the segmented airway tree into branches, but rather on a sub-tree spanned by the labelled branches, as explained in Section 4.2.4. This enables the algorithm to tackle structural noise such as false or missing branches, as opposed to methods that work only with the originally segmented branches. The hierarchical implementation makes the algorithm sufficiently fast to be of practical use.

A thorough evaluation of the labelling algorithm is made on a set of 80 segmented airway trees from 40 subjects scanned at two different time-points.

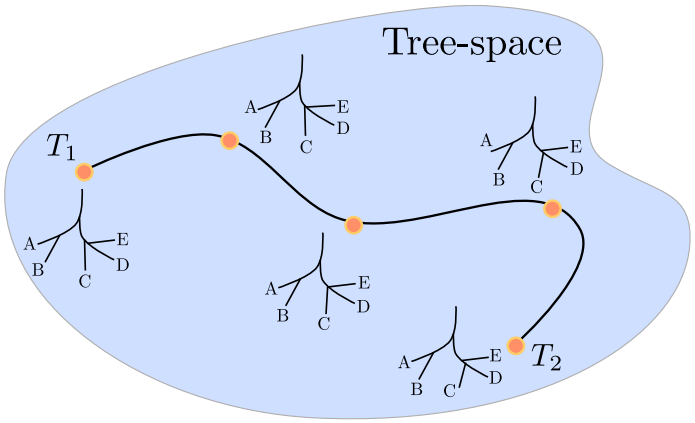


Figure 4.2: Since tree-space is a path connected space, any two trees are joined by a path in tree-space, corresponding to a tree deformation along which tree topology and branch geometry changes. The geodesic distance between two trees is the length of the shortest path connecting them.

Each airway tree was labelled by 3 clinical experts. There are subjects without COPD and subjects with different stages of COPD, ranging from mild to severe. The evaluation includes a comparison of accuracy and reproducibility of the algorithm to that of the clinical experts, as well as robustness of accuracy and reproducibility to disease severity. Moreover, reproducibility and robustness of reproducibility to COPD severity is also tested on a large data set from a longitudinal lung cancer screening trial using 8724 CT images from 1900 individuals.

A preliminary version of the work presented here appeared in Feragen et al. 2012. In comparison with the earlier paper, changes have been made to the algorithm, making the hierarchy less sensitive to missing RUL branches as detailed in Section 4.2.5. The exposition has been extended to give a far more comprehensive explanation of the tree-space methodology used. Finally, the experimental validation has been significantly extended. First, our labelled data set has been doubled in size and is now manually annotated by three clinical experts. Second, an evaluation of reproducibility on a large longitudinal study has been conducted, as well as a statistical analysis showing that correlation between increased COPD severity and decreased labelling reproducibility is due to segmentation problems in patients with COPD.

4.2 Methodology: Branch labelling

The airway branch labels illustrated in Figure 4.6 correspond to the division of the lung into compartments: LMB and RMB lead to the left and right lungs; LUL, RUL, L4+5, R4+5, LLB, and RLL lead to the different lobes; and R1-R10, L1-L10 lead to the segments, with up to 10 segments in each

lung. In addition, a number of intermediate branch names appear in the literature, whose presence in a particular anatomical airway tree depends on its topology. If the locations of all segment branch labels are known, along with the airway tree structure, then it is straight-forward to reconstruct the remaining branch labels. In this sense, a *leaf-labelled airway tree*, where the leaf labels are segment labels, is equivalent to a *labelled airway tree*.

4.2.1 Input data

The input to the labelling algorithm is a connected centreline tree extracted from an airway tree segmentation, divided into branches by bifurcation points. The airway segmentation, centreline extraction and bifurcation detection algorithms used in our experiments are detailed in Section 4.3.1.

Based on the extracted airway centrelines, each branch is represented by 6 landmark points sampled equidistantly along the centreline, translated so that the first landmark point is placed at the origin. Thus, ignoring the first origin landmark point, each branch e is represented by a vector $x_e \in (\mathbb{R}^3)^5 = \mathbb{R}^{15}$. Each airway tree is normalized by the person's height as an isotropic scaling parameter. The person's height was chosen over alternative normalization parameters such as lung volume because height is unaffected by disease.

4.2.2 The labelling algorithm: An overview

The general goal of the labelling algorithm is, for an arbitrary unlabelled airway tree T, to optimally assign the set of segment labels $\{L1, ..., L10, R1, ..., R10\}$, corresponding to the 20 segment bronchi, to branches in the centreline tree. As outlined in Algorithm 1, the basic labelling algorithm contains a labelling suggestion step and a labelling evaluation step, after which an optimal labelling is selected. In practice, for the sake of computational efficiency, this algorithm is repeated in a hierarchical fashion, assigning labels from the top and downwards. This is detailed in Sec. 4.2.5.

Algorithm 1 Overview of the basic labelling algorithm

1: **Input:** Unlabelled tree T
2: **Input:** Training set of labelled trees $\{T_i | i \in I\}$
3: Generate a set $\mathcal{L}$ of suggested labellings $L \in \mathcal{L}$
4: **for** suggested labellings $L \in \mathcal{L}$ **do**
5: $T_L \leftarrow T$ with suggested labelling L
6: Compute distances $\{d(T_L, T_i) | i \in I\}$
7: Compute $f_L = \sum_{i \in I} d(T_L, T_i)$
8: **end for**
9: **Output:** Labelled tree $T_L = \text{argmin}_{L \in \mathcal{L}} f_L$

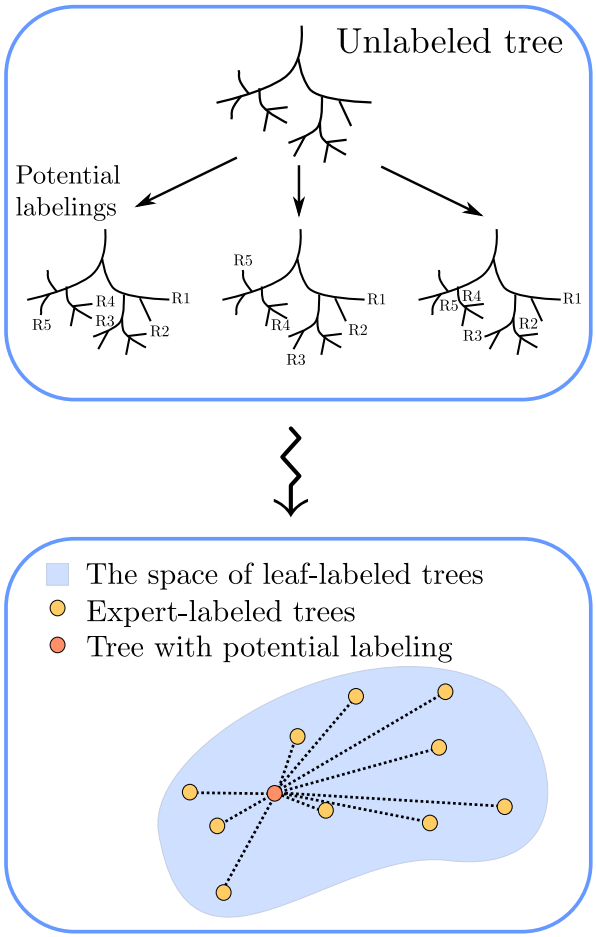


Figure 4.3: The labelling algorithm contains two steps repeated in a hierarchical fashion: Generation of suggested labellings (top) and evaluation of suggested labellings based on tree-space distances to expert-labelled training trees (bottom).

Algorithm 1 contains a label suggestion step (line 3) and a label selection step (line 9) as illustrated in Figure 4.3. In the label suggestion step, a series of potential label configurations are suggested. This is explained in detail in Section 4.2.5. In the labelling step, the optimal branch label assignment is selected as the configuration that minimizes the sum of distances to manually labelled trees from a training set. The distance used is the geodesic distance in the space of leaf-labelled trees, as detailed in Section 4.2.6 below.

4.2.3 Trees

By *tree* we mean a rooted tree, defined as a triple $T = (V, E, r)$ where V is a finite set of vertices, $E \subset V \times V$ is a set of edges so that there are no cycles

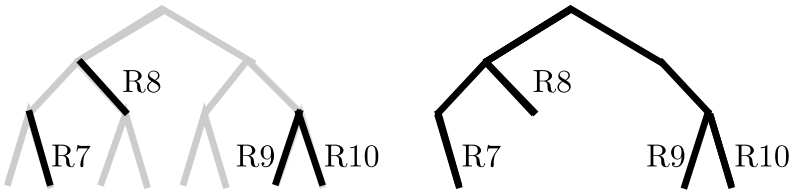


Figure 4.4: From a configuration of leaf labels we extract the sub-tree spanned by the labels and prune off the rest, obtaining *the sub-tree spanned by the labels*, a leaf-labelled tree which can be compared to the training trees.

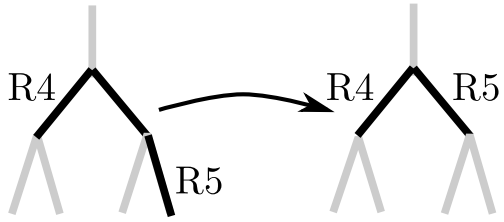


Figure 4.5: Each assigned label is backtraced to the branch closest to the root that is not already part of the sub-tree spanned by the other labels.

in the corresponding graph, and r is a designated root vertex. Anatomical or biological transportation systems often have a natural source node which can be used as a root. For airway trees, the trachea provides a natural and easy-to-identify root branch. Given any edge $e \in E$, any other edge $\tilde{e} \in E$ which sits on the path through the tree from e to the root is said to be *above* e. If $\tilde{e}$ is above e, then we say that e is *below* $\tilde{e}$.

A *labelling* of T is a map $L\colon X \to E$, which assigns unique labels from a label set X to some but not necessarily all edges. In this paper, we are particularly interested in *leaf-labelled trees*. A *leaf* in T is an edge which does not have any other edges below it. A leaf-labelled tree on the leaf label set X is a tree endowed with a bijective labelling $L\colon X \to E_l$, where $E_l \subset E$ are the leaf edges in T. In particular, $|X|$ must equal the number of leaves in T.

4.2.4 From labelled airway trees to leaf-labelled trees

Given segmented airway trees, we wish to extract leaf-labelled sub-trees in such a way that particular sets of branches play the roles of leaves; for instance, the fixed set of leaf labels $\{L1, ..., L10, R1, ..., R10\}$. However, segmented airway trees have variable size and usually, many branches are detected below the segment level. In order to study airway trees using a framework for leaf-labelled trees we define, given any labelling $L\colon X \to E$, the *sub-tree spanned by the labels* as the tree obtained by removing all edges in the tree which are not found on the path from the root to an edge labelled by L, as in Figure 4.4. Consecutive edges which are joined by a vertex of order 2 will be concatenated,

as is the case with parent branch of $R9$ and $R10$ in Figure 4.4. When two edges e_1 and e_2 are concatenated into an edge e, the shape vector x_e will be recomputed from the concatenation of the branch centrelines corresponding to e_1 and e_2. After labelling, each label is backtraced through the path to the root, as in Figure 4.5. We only consider *admissible* label configurations, defined as labellings where the leaf labels will all be attached to *leaves* in the sub-tree spanned by the labels. This is equivalent to excluding labellings where two leaf labels are assigned to branches on the same path to the root.

For a tree T with a labelling $L\colon X \to E$, we denote by T_L the sub-tree spanned by labels assigned by L.

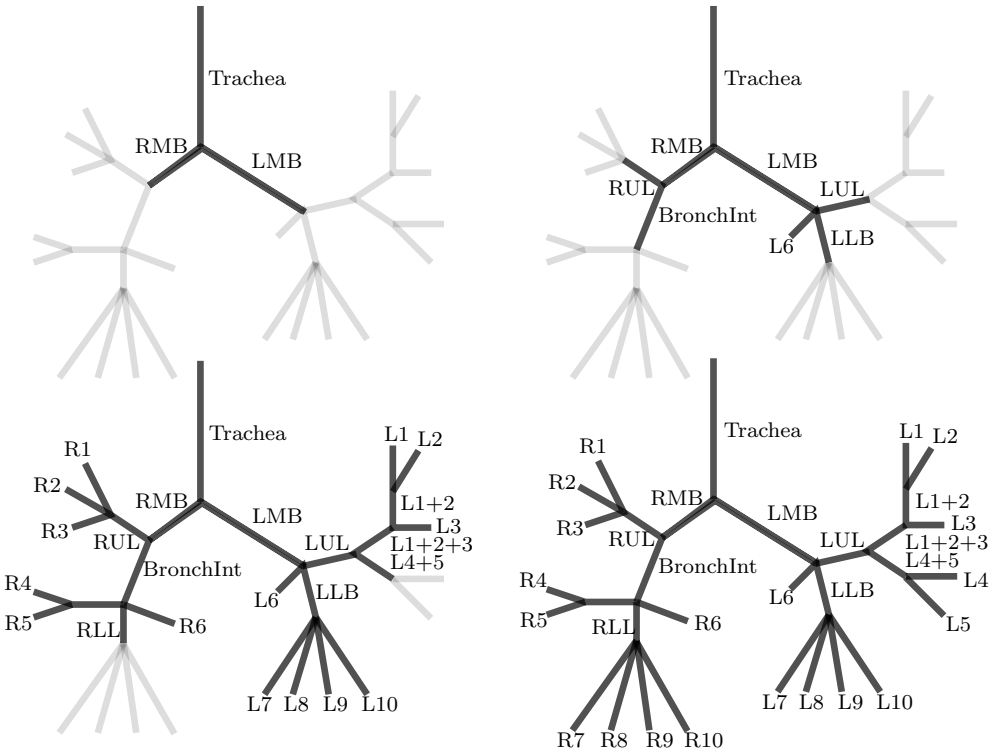


Figure 4.6: Hierarchical labelling: In each step, search through a sub-tree for an optimal alignment of a subset of labels, obtaining a leaf-labelled sub-tree of the segmented airway tree similar to the trees shown in black. The real tree topology may differ; the figure only illustrates the stepwise hierarchy.

4.2.5 A hierarchical algorithm

Ideally, we would search through the whole airway tree T, test all admissible configurations T_L of the 20 segment leaf labels and select the one that optimizes line 9 in Algorithm 1. However, for an airway tree with as few as 100 branches, the search space size is on the order of 100^{20}, which is too

Algorithm 2 For computational speed, the labelling is split into a set of hierarchical sub-tree labelling steps.

1: Label the first branch in the airway tree as the trachea.
2: Search 3 generations below the trachea for the optimal configurations of the label set $X_1 = \{\text{LMB, RMB}\}$.
3: Search 2 generations below the RMB for the optimal configurations of the label set $X_2 = \{\text{RUL, BronchInt }\}$.
4: Search 2 generations below the LMB for the optimal configurations of the label set $X_3 = \{\text{L6, LLB, LUL}\}$.
5: Search 2 or 3 generations in "anything but the BronchInt tree" for optimal configurations of $X_4 = \{\text{R1, R2, R3}\}$.
6: Search 2 generations below the BronchInt for optimal configurations of $X_5 = \{\text{R4, R5, RLL, R6}\}$.
7: Search 2 generations below the LLB for optimal configurations of $X_6 = \{\text{L7, L8, L9, L10}\}$.
8: Search 3 generations below the LUL for optimal configurations of $X_7 = \{\text{L1, L2, L3, L4+5}\}$.
9: Search 3 generations below the RLL for optimal configurations of $X_8 = \{\text{R7, R8, R9, R10}\}$.
10: Search 2 generations below the L4+5 for optimal configurations of $X_9 = \{\text{L4, L5}\}$.

large to handle. In order to ensure computational feasibility, we choose a hierarchical sub-tree approach, where labels of different generations are added subsequently, as explained in Figure 4.6 and Algorithm 2. Here, more shallow branches are treated as leaves in the first steps of the algorithm, which works its way down to the segments. In each step of the hierarchical label placement, the optimal branches for the given set of labels is selected.

In each step of the hierarchical labelling, a specific set of descendants are assigned to a specific already labelled branch. For instance, in line 4 of the algorithm, the LMB branch has already been assigned, the sub-tree rooted at LMB is extracted, and the algorithm attempts to assign leaf labels {L6, LLB, LUL} in any possible configuration spanning two generations below the LMB. The reason for searching 3 generations in some cases and 2 in others is a trade-off between having enough space in the tree to assign all branches in a given step, and having a small enough tree to limit the number of possible configurations for the sake of computational speed. This trade-off depends on the number of labels assigned in a step as well as the likelihood of having higher order nodes in that particular sub-tree, as higher order nodes give more branches per generation.

The step in line 5 is different from the others. Since the RUL branch is not always present, it is problematic to root a hierarchy sub-tree at the RUL as is done in Feragen et al. 2012. Thus, rather than searching the tree below the

RUL branch, assigned in line 3, we search the tree obtained from the sub-tree rooted at the parent of RUL by removing the sub-tree rooted at BronchInt.

The choice of "leaves" used at the different steps in the hierarchy was made in order to minimize the number of "leaves" used while only using "leaves" that actually appear in as many people as possible. In a different application we would recommend a similar strategy.

4.2.6 Tree-space and tree-space distances

The tree-to-tree distances used in this paper are measured in a *tree-space*. This tree-space is a straight-forward generalization of the phylogenetic tree-space from Billera et al. 2001, where single-dimensional shape vectors on the branches have been generalized to multi-dimensional ones. Each point in tree-space is a leaf-labelled tree, with leaves labelled by some fixed set X, for instance $X = \{L1, ..., L10, R1, ..., R10\}$. Tree-space is a *path connected space*, which means that any two trees can be joined by a path in tree-space. Moving along such a path corresponds to deforming the trees, as in Figure 4.2. Moreover, in this tree-space there will always be a *shortest* path, called a *geodesic*, joining any given two trees. The length of the geodesic defines a distance between the two trees (Billera et al. 2001), called the *geodesic distance*. In this way, we obtain a metric distance measure on tree-space.

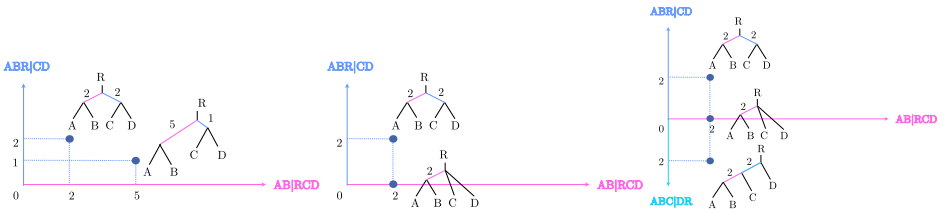
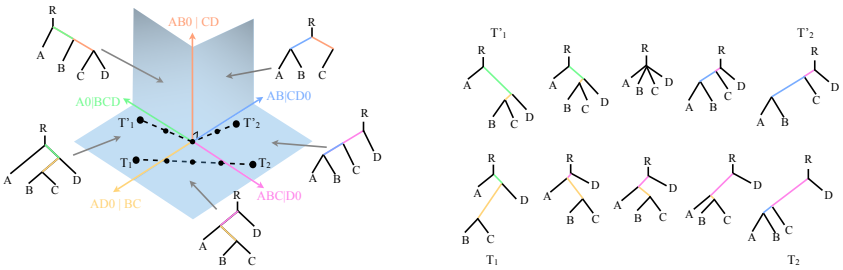


Figure 4.7: Tree-space is a union of *orthants*, each corresponding to a specific leaf-labelled tree topology. (a) Different points in the orthant are trees with identical topology but different shapes. (b) Points at the boundary of an orthant are points where one edge is described by a zero vector. Geometrically, that edge has been contracted. (c) Orthants with different tree topologies meet at the boundaries where the contracted edges give rise to new, identical tree topologies.

Each set of trees having a given topology forms an *orthant*, as shown in Figure 4.7 and 4.8, which is a lower-dimensional Euclidean space (or, in the case where edges are described by edge length, a positive orthant of a Euclidean space)[1]. The algorithm (Owen and Provan 2011) for computing the

[1]Formally, an orthant is the part of Euclidean space where all coordinates are non-



(a) (b)

Figure 4.8: (a) Tree-space is a proper subset of the Euclidean space $(\mathbb{R}^{15})^S$, and the distance between two trees T_1 and T_2 is the length (measured in the Euclidean space) of the geodesic, or shortest path, in tree-space from T_1 to T_2. Note that the geodesic from T_1' to T_2' is *not* a straight line, giving different topological transitions throughout the two paths connecting T_1 to T_2, and T_1' to T_2'. This is illustrated in (b), where trees are sampled along the two geodesic paths. *We illustrate the tree-space using edge length for edges rather than their 3D shape; this is done for illustrative purposes only. The same behaviour carries over to edges with shape-vector attributes. Furthermore, the 5 axes depicted above each correspond to their own dimension, and have only been embedded into $\mathbb{R}^3$ for illustrative purposes.*

geodesic distance between two trees works by recursively determining the sequence of orthants containing the geodesic as follows. If the trees are in the same orthant (i.e. if they have the same topology), the algorithm terminates and returns the Euclidean distance between the two trees in the orthant. Otherwise, if the trees are in different orthants, the algorithm starts by computing an initial path connecting the two trees, which goes straight from the first tree, to the origin, and back to the second tree. The algorithm checks if this is the geodesic by looking for an orthant such that modifying the current path to go through this orthant instead of the origin gives a shorter path, i.e., whether this orthant provides a "short-cut" which avoids the origin. If such an orthant exists, we add it to our sequence and calculate the geodesic through the three orthants. Again, we check if this is the overall geodesic by checking each point where the path changes orthants for a new orthant to add into the orthant sequence, such that going through this new orthant will give a shorter path. The new geodesic through the expanded orthant sequence is computed, and this process is repeated until no more orthants can be added. The length of the geodesic through the resulting orthant sequence is the geodesic through

negative. When edges are described by their length, as with phylogenetic trees, tree-space orthants are precisely such Euclidean orthants; we call them *positive orthants*. In our paper, branches can have negative 3D coordinates, making tree-space orthants larger than Euclidean orthants. To keep the terminology consistent with phylogenetic tree-space papers, we use the word "orthant" for these larger sub-spaces of tree-space.

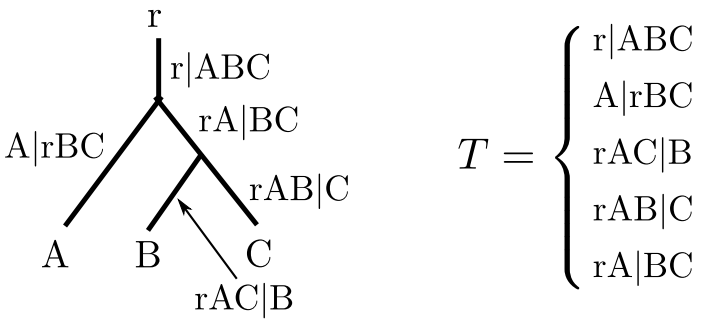


Figure 4.9: Tree edges are topologically identified with partitions of the leaf label set X, and a tree topology is characterized uniquely by the partitions that define its edges.

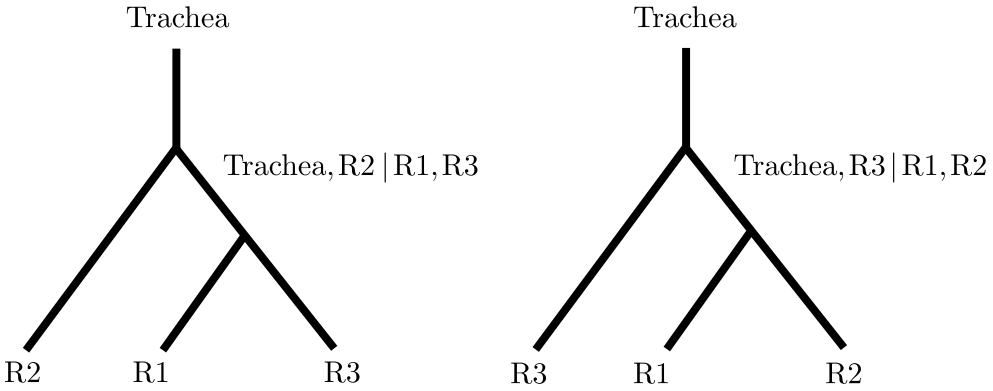


Figure 4.10: Certain pairs of label set partitions represent edges that cannot exist in a tree simultaneously. An example for the leaf label set $X = \{$R1, R2, R3$\}$ is an edge that splits $\{$R1, R2$\}$ off from the rest of the tree and an edge that splits $\{$R1, R3$\}$ off.

tree-space. See Owen and Provan 2011 for further details and code.[2]

Another way to understand the geodesic distance between trees comes from the fact that tree-space is naturally embedded as a subset of a high-dimensional Euclidean space, which can be seen as follows:

Each edge in the leaf-labelled tree can be identified topologically by a partition of X into the leaves below the edge, and the remaining leaves (including the root), as in Figure 4.9. Let S denote the set of all possible partitions of X. Any leaf-labelled tree topology with leaf label set X corresponds to a binary vector $\{0, 1\}^S$, where each coordinate $s \in S$ that is set to 1 indicates that the particular label set bipartition s identifies an edge in the tree topology. Adding shape to the picture, a tree will uniquely correspond to a vector in

[2]Code available from: `http://vm1.cas.unc.edu/stat-or/webspace/miscellaneous/provan/treespace/`.

$(\mathbb{R}^{15})^S$, where each consecutive set of 15 coordinates corresponds to a possible partition s of X. If the edge associated with that partition appears in the tree, then those 15 coordinates will be its branch vector, and otherwise they are all 0. Certain edges can never appear in a tree together. An example is shown in Figure 4.10, where an edge that splits $\{R1, R2\}$ off from the rest of the tree and an edge that splits $\{R1, R3\}$ off from the rest of the tree cannot possibly appear in the same tree. This means that *tree-space* is not all of $(\mathbb{R}^{15})^S$, but consists precisely of those vectors in $(\mathbb{R}^{15})^S$ that correspond to trees, and tree-space is a proper subset of Euclidean space. It can be shown that the geodesic distance $d(T, T')$ between two trees T and T' defined in Billera et al. 2001 coincides with the length of the shortest path between T and T' that remains fully within this restricted subspace, length being measured in the ambient Euclidean space using the Euclidean metric. An analytic formula for this distance d does not exist, but it can be computed recursively in polynomial time (Owen and Provan 2011).

Tree-space geodesics are further illustrated in Figure 4.8, where a concrete example is given of two geodesics for which the endpoint trees (T_1, T_2) and (T_1', T_2') have identical topology, but the topological transitions taking place throughout the geodesic from T_1 to T_2 are not the same as those taking place in the geodesic from T_1' to T_2'.

In terms of the ambient Euclidean space, the geodesic connecting T_1' and T_2' is not a straight line, because tree-space does not fill out the whole ambient Euclidean space. A straight line path from T_1' to T_2' in $(\mathbb{R}^{15})^S$ would have to pass through the orthant formed from the partitions/axes $AR|BCD$ and $ACD|BR$, so some intermediate trees would have to simultaneously contain i) an edge that splits the labels B, C and D off from A and the root, as well as ii) an edge that splits the edges A, C and D off from B and the root. Clearly, the partitions $\{A, R\}$ and $\{B, R\}$ cannot both happen in the same tree (they are incompatible); hence that orthant does not exist in tree-space.

4.2.7 The detailed labelling algorithm

Based on the previous sections, we now give a more detailed version of the labelling algorithm. For each step in the hierarchical labelling in Algorithm 2, with a fixed leaf label set X, we apply Algorithm 3 as follows:

Given a set of expert-labelled training trees T_i, $i \in I$, replace each and every one of them by the sub-tree spanned by the labels X. Similarly, given an unlabelled airway tree T and a proposed leaf-labelling L of T with leaf-label set X, first extract the sub-tree T_L of T spanned by the labels.

For each leaf-labelled tree T_i in our training set and each T_L, we compute the geodesic distance $d(T_i, T_L)$ between the trees T_i and T_L in the tree-space defined in Section 4.2.6. We extract a labelling of T with label set X by

choosing the labelled tree T_{labelled} among the T_L that satisfies:

$$T_{\text{labelled}} = \arg\min_{T_L} \sum_{i \in I} d(T_i, T_L). \tag{4.1}$$

Algorithm 3 The detailed labelling algorithm

1: **Input:** Unlabelled tree T
2: **Input:** Training set of labelled trees $\{T_i | i \in I\}$
3: **for** $i \in I$ **do**
4: $T_i \leftarrow$ sub-tree of T_i spanned by the labels
5: **end for**
6: Generate a set $\mathcal{L}$ of suggested labellings of T, denoted $L \in \mathcal{L}$.
7: **for** suggested labellings $L \in \mathcal{L}$ **do**
8: $T_L =$ sub-tree of T spanned by the labels assigned by L
9: Compute distances $\{d(T_L, T_i) | i \in I\}$
10: Compute $f_L = \sum_{i \in I} d(T_L, T_i)$
11: **end for**
12: **Output:** Labelled tree $T_L = \text{argmin}_{L \in \mathcal{L}} f_L$

4.3 Experimental results

We evaluate three different aspects of labelling performance. First, we evaluate labelling accuracy, defined as the ability to assign labels to the same branches as clinical experts. Second, we evaluate labelling reproducibility, defined as the ability to assign labels to the same branches in airway trees extracted from repeated scans of the same subject. Third, we evaluate the dependence of accuracy and reproducibility on COPD diagnosis and severity. In all three aspects, the performance of the algorithm is compared to the average performance of clinical experts.

4.3.1 Data

The data used in the experiments comes from the Danish Lung Cancer Screening Trial (Pedersen et al. 2009). It consists of low-dose (120 kV and 40 mAs) pulmonary CT scans and lung function measurements. The scans were obtained from a Multi Detector CT scanner (16 rows Philips Mx 8000), reconstructed using a hard algorithm (kernel D) with a resolution of approximately 0.78mm × 0.78mm × 1mm. The lung function measurements, used to determine COPD severity, were performed using a computerized system (Spirotrac IV, Vitalograph) according to recommendations by the European Respiratory Society (Miller et al. 2005) without the use of bronchodilation.

The airway lumen surface was extracted from the images using the locally optimal path approach of Lo et al. 2009 and then refined using the optimal

surface approach of Chapter 2. Afterwards centrelines were computed by front propagation within the refined lumen surface as described in Lo et al. 2012. The resulting centrelines were disconnected in bifurcations regions and so Dijkstra's algorithm was used to connect them along shortest paths within an inverted distance transform of the refined lumen surface.

A data set of 80 airway tree centrelines from 40 subjects scanned at two time-points with intervals of 5 years, were manually assigned segment labels $L1$ - $L10$ and $R1$ - $R10$ by two experts in pulmonary medicine (LHT and AD) and one in radiology (MMWW). The labels were assigned according to Netter 1989 and Feneis 1995. The experts were allowed to assign the same label to multiple branches in cases where they were unsure. The manual labelling was done using in-house developed software, simultaneously showing the segmented airway and centreline, which can be rotated, panned and zoomed, as well as a CT cross-section perpendicular to and centred on any given point of the airway. The remaining labels seen in Figure 4.6 were deduced from the segment labels.

COPD severity was defined according to the GOLD 2013, from the averaged lung function at both time-points. Out of the 40 subjects, there were 9 subjects with no airflow limitation, and 11 with mild, 11 with moderate, and 9 with severe COPD. We will denote the groups as GOLD 0, GOLD 1, GOLD 2, and GOLD 3, respectively.

The algorithm was further tested in a large longitudinal data set including all the subjects from the Danish Lung Cancer Screening Trial (Pedersen et al. 2009) who had at least two usable scans. For this data set lung CT image registration (Gorbunova et al. 2012) was used to automatically determine reproducibility, as described in Section 4.3.5, and so it was important that the images could be registered well. A scan was therefore deemed non-usable if the lungs were not entirely contained within the image or if bowel air was erroneously included within lung segmentations (lung segmentation method and manual validation are described in Ashraf et al. 2011). This resulted in the inclusion of 1900 subjects of which, based on average lung function measurements over all time-points, 975 belonged to GOLD 0, 495 belonged to GOLD 1, 391 belonged to GOLD 2, and 38 belonged to GOLD 3. There was a single subject with very severe COPD (what would otherwise be GOLD 4), which was included in the GOLD 3 group. Each of these subjects had an average of 4.6 ± 0.7 usable scans approximately evenly distributed over a period of 5 years.

4.3.2 Implementation

The labelling algorithm was implemented in MATLAB, using tree distance computations implemented in Java[3]. For the annotated dataset, the airway

[3]Code available from `http://vm1.cas.unc.edu/stat-or/webspace/miscellaneous/provan/treespace/`.

	Accuracy		Reproducibility		Labelled airways	
Label	Automatic	Expert	Automatic	Expert	Automatic	Expert
R1	89.2%	87.9%	95%	91.1%	80.0	80.0
R2	87.5%	84.2%	97.5%	90%	80.0	80.0
R3	87.6%	87.2%	97.5%	86.7%	80.0	80.0
R4	90.4%	88.7%	92.5%	86.4%	80.0	79.3
R5	86.9%	84.3%	90%	82.1%	80.0	79.0
R6	91.8%	93.3%	97.5%	91.5%	80.0	80.0
R7	77.7%	79.0%	84.6%	85.2%	76.0	79.7
R8	72.8%	75.8%	69.2%	79.2%	76.0	80.0
R9	63.2%	67.5%	53.8%	55.7%	76.0	80.0
R10	59.1%	64.3%	51.3%	55.8%	76.0	80.0
L1	64.0%	54.4%	75%	58.9%	79.9	79.7
L2	65.8%	60.1%	70%	62.2%	79.9	79.7
L3	66.0%	59.3%	75%	64.2%	79.9	80.0
L4	69.4%	78.3%	72.5%	80.7%	79.9	80.0
L5	73.2%	84.6%	70%	86.7%	79.9	80.0
L6	99.6%	99.2%	100%	99.2%	80.0	80.0
L7	62.8%	53.8%	82.5%	63.6%	80.0	77.3
L8	54.9%	48.6%	87.5%	57.1%	80.0	79.7
L9	53.4%	58.3%	72.5%	62.2%	80.0	80.0
L10	58.3%	57.1%	80%	62.5%	80.0	80.0
Trachea	100.0%	100.0%	100.0%	100.0%	80.0	80.0
LMB	100.0%	100%	100%	100%	80.0	80.0
LUL	100.0%	100%	97.5%	97.5%	80.0	80.0
LB1+2	62.9%	50%	70.6%	65.3%	61.8	57.0
LB4+5	92.3%	89.6%	95%	91.7%	79.9	78.7
LLB	99.2%	98.3%	100%	98.3%	80.0	79.3
RMB	100.0%	100%	100%	100%	80.0	80.0
RUL	97.4%	95%	100%	100%	78.0	76.0
BronchInt	99.9%	99.7%	100%	100%	80.0	80.0
RB4+5	95.8%	95.4%	95%	95%	80.0	78.0
RLL	93.3%	96.7%	95%	96.7%	80.0	79.7
LB1+2+3	92.0%	81.7%	94.6%	89.9%	74.9	69.0
Segmental average	73.7 ±4.8%	73.3 ±9.7%	80.8 ±16.3%	75.0 ±10.2%	79.2	79.7
Total average	81.5 ±3.5%	80.4 ±7.1%	86.4 ±13.1%	82.4 ±7.2%	78.7	78.5

Table 4.1: The mean accuracy of the algorithm was computed from 10 repetitions of 10-fold cross validation, and the mean accuracy of an expert was averaged over all three pairs of experts. The third and fourth columns contain the mean reproducibility of the algorithm and an expert, respectively. The mean number of airways in which a given label was assigned, was averaged over 10 cross-validation runs or three experts, respectively.

trees had 181 branches on average, and the whole labelling took roughly 10 minutes per tree running on a laptop with a single 2.40 GHz processor using no more than 3 GB RAM per labelling.

Labelling experiments on the annotated dataset were performed using 10-fold cross validation, where both scans of any individual were always contained in the same fold. Thus, for each test set fold of 8 airway trees from 4 patients, the training set was made up of 72 airway trees from 36 patients. Each tree was labelled separately by the three medical experts, but sometimes the medical experts would, when in doubt, place the same label on two different branches. In these cases two leaf-labelled training trees would be generated, one for each option. In other cases, some labels were not assigned by the expert, in which case the corresponding training sub-tree would not be generated. This resulted in 231-280 training airway trees from the 80 scans, with different numbers at different steps of the hierarchy.

4.3.3 Labelling results: Accuracy

The accuracy of the automatic labelling, defined as its ability to agree with a clinical expert, was assessed using 10 labelling runs of 10-fold cross validation with randomized folds. For each airway, the average success rate was computed out of the number of labels assigned by either algorithm or expert (meaning that if neither the algorithm nor the expert assign a given label, then this label does not contribute to the success rate of that airway tree at all).

It is not obvious how labelling accuracy should be assessed. In some cases where experts were not certain, or judged that an anatomical branch had been split into two branches by the segmentation algorithm, they would assign the same segment label to multiple branches (the three experts did this in 34, 12 and 5 of the 80 airway trees, respectively). In other cases branches were missing, either anatomically or from the segmentation, so that some labels were not assigned (the three experts did this in 10, 1 and 4 of the 80 airway trees, respectively). Missing label assignments happened both in expert and automatic labellings. Sometimes branches would be missing in the airway tree, making label assignments impossible. Other times, the topology of the airway made certain non-segment labels impossible. Thus, Table 4.1 contains average counts for how many times labels were assigned by the algorithm and the experts, respectively.

In order to fairly assess all cases, we gave, for the i^{th} airway tree and each label $x \in X = \{\text{L1, ..., L10, R1, ..., R10}\}$, the assignment by method M_1 a correctness percentage $s_i(x, M_1, M_2)$ with respect to method M_2. M_1 could be either an expert or the automatic labelling, and M_2 was always an expert. The correctness percentage was defined as follows: In the i^{th} tree T_i, let $M_1(T_i, x)$ denote the set of branches assigned label x by method M_1 and $M_2(T_i, x)$ the set of branches assigned label x by method M_2. Define the

correctness $s_i(x, M_1, M_2)$ of label x using method M_1 with respect to method M_2 in the i^{th} airway tree as:

$$s_i(x, M_1, M_2) = 100 \cdot \frac{2 \cdot |M_1(T_i, x) \cap M_2(T_i, x)|}{|M_1(T_i, x)| + |M_2(T_i, x)|}\%,$$

We assume cases where the label was not assigned by either method, that is $|M_1(T_i, x)| = |M_2(T_i, x)| = 0$, to be due to missing branches and thus leave them out of the total summary shown in Table 4.1.

On average, the automatic labelling agreement with an expert was $73.7 \pm 4.8\%$ on the segment branches, and $81.5 \pm 3.5\%$ overall. This is not significantly different from the average expert agreement with an expert, which was $73.3 \pm 9.7\%$ on the segment labels, and $80.4 \pm 7.1\%$ overall ($p = 0.94$ and $p = 0.77$ in Mann-Whitney U-tests).

Figure 4.11 shows labelling accuracy stratified by COPD severity. Spearman's correlation test shows no significant correlation between the average agreement with an expert and the presence and severity of COPD ($\rho = -0.18$, $p = 0.11$ on all labels, $\rho = -0.20$, $p = 0.069$ on segment labels). Similarly, there is no correlation between the average agreement between experts, and presence and severity of COPD ($\rho = -0.12$, $p = 0.45$ on all labels, $\rho = -0.085$, $p = 0.60$ on segment labels).

4.3.4 Labelling results: Reproducibility of expert and automatic labelling

In order to test reproducibility of the expert and automatically assigned labels, the two CT scans of each subject were registered using deformable image registration as described in Gorbunova et al. 2012, and the labelled airway branches were manually investigated for possible matches in the resulting common coordinate system. Let T_i^1 and T_i^2 be two trees corresponding to the i^{th} subject's airway at time-points 1 and 2, and let $M(T_i^1, x) \subseteq E_i^1$ and $M(T_i^2, x) \subseteq E_i^2$ be sets of branches assigned label x by the method M in T_i^1 and T_i^2, respectively. Denote by $R\big(M(T_i^1, x), M(T_i^2, x)\big)$ the matched subset of these branches.

We define the reproducibility of label x using the method M within the trees T_i^1 and T_i^2 of subject i

$$r_i(x, M, T_i^1, T_i^2) = 100 \cdot \frac{2 \cdot |R\big(M(T_i^1, x), M(T_i^2, x)\big)|}{|M(T_i^1, x)| + |M(T_i^2, x)|}\%.$$

To avoid evaluating effects of missing branches due to segmentation problems, cases where the label was not assigned in either time-point, that is $|M(T_i^1, x)| = |M(T_i^2, x)| = 0$ are left out of the total summary. The same holds for cases where only one time-point was labelled with x, which without loss of generality can be assumed to be T_i^1, but only if matching branches did not exist in the other time-point, that is $R(M(T_i^1, x), T_i^2) = \emptyset$. Table 4.1 shows a summary of the results.

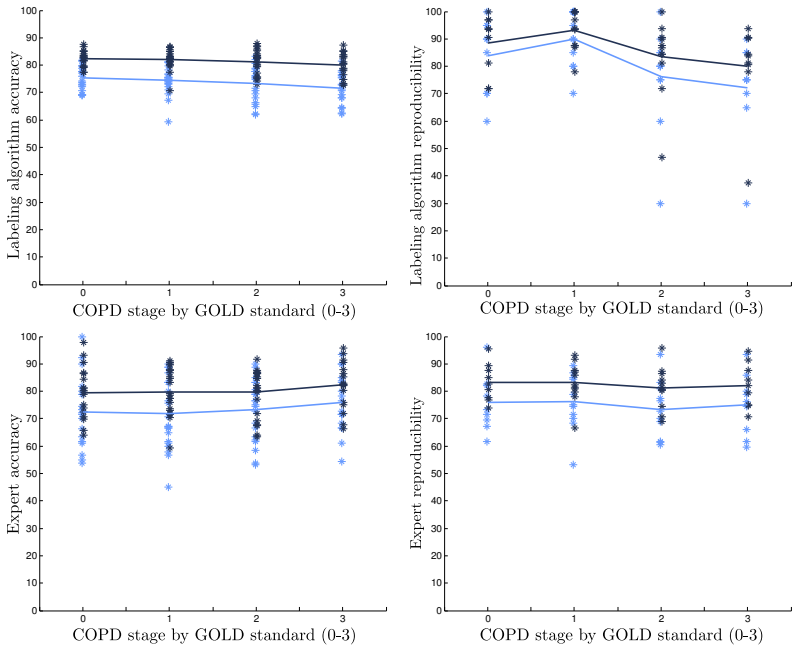


Figure 4.11: Average labelling accuracy (left) and reproducibility (right) for segment labels (light blue) and all labels (dark blue), stratified by COPD severity for the algorithm (top) and experts (bottom). Each $*$ corresponds to accuracy/reproducibility for one subject. The lines interpolate mean accuracies and reproducibilities for each GOLD group.

On average, the reproducibility of the automatic labelling was $80.8 \pm 16.3\%$ on the segment labels, and $86.4 \pm 13.1\%$ overall, which is significantly better than the reproducibility of the experts, which was $75.1 \pm 14.8\%$ on the segment labels, and $82.4 \pm 10.6\%$ overall ($p = 0.021$ and $p = 0.022$ in Mann-Whitney U-tests).

Figure 4.11 shows labelling reproducibility stratified by COPD severity. Spearman's correlation test shows significant correlation between reproducibility of the automatic approach and severity of COPD ($\rho = -0.34$, $p = 0.031$ on all labels; $\rho = -0.36$, $p = 0.024$ on segment labels). Spearman's correlation test shows, however, no significant correlation between the average reproducibility of the expert labelling and the severity of COPD ($\rho = -0.085$, $p = 0.604$ on all labels; $\rho = -0.049$, $p = 0.764$ on segment labels).

4.3.5 Labelling results: Reproducibility on large longitudinal data set

Reproducibility of the automatic approach on the larger data set was tested by labelling scans using the manually labelled airway trees as a training set. All scans of each subject were registered (Gorbunova et al. 2012) and branches

appearing in multiple images were matched, in a similar fashion to what was described in the previous section. However, rather than manually detecting matched branches, which would be very time consuming for a data set of this size (8724 trees), matching was done automatically. The details of this automatic matching approach have previously been published in Petersen et al. 2011a.

It is not obvious how labelling reproducibility should be defined in a subject with more than two time-points. As an example, consider a case where three out of five time-points agree on one assignment of the label x, and the remaining two time-points agree on another. Taking such cases into account, we define reproducibility of assigning the label x in terms of percentage agreement with the majority labelling (if two labellings are both majority, one of them is just selected).

To define reproducibility analytically, let $T_i^1, \ldots, T_i^n$ denote the airway trees of subject i at n different time-points, and let $M(T_i^1, x), \ldots M(T_i^n, x)$ be the branches assigned label x by method M in each time-point. In addition let the set of time-points where the assignment of the label x matches the assignment in time-point j be:

$$N_i(M, x, j) = \Big\{k \in \{1, \ldots, n\} \mid R\big(M(T_i^j, x), M(T_i^k, x)\big) \neq \emptyset\Big\}.$$

A time point where the majority labelling occurs is then given by:

$$j_{\max}(x, M) = \underset{j \in \{1,\ldots,n\}}{\operatorname{argmax}} |N_i(x, j, M)|.$$

To avoid evaluating effects of missing branches due to segmentation problems, time-points $k \in \{1, \ldots n\}$ are left out if x has not been assigned, that is $M(T_i^k, x) = \emptyset$, and none of the other branches match the branch labelled with x in a majority labelled time-point, that is $R\big(M(T_i^{j_{\max(x,M)}}, x), T_i^k\big) = \emptyset$. The remaining time-points are denoted by $P_i(x, M)$:

$$P_i(x, M) = \Big\{k \in \{1, \ldots, n\} \mid R\big(M(T_i^{j_{\max(x,M)}}, x), T_i^k\big) \neq \emptyset \text{ or } M(T_i^k, x) \neq \emptyset\Big\}.$$

We then define reproducibility of a label x in subject i by method M as the percentage of time-points agreeing with the majority labelling out of the total amount of included time-points:

$$r_i(x, M) = 100 \cdot \frac{|N_i(x, j_{\max(x,M)}, M)|}{|P_i(x, M)|}\%,$$

where $|N_i(M, x, j_{\max(x,M)})| > 1$, otherwise $r_i(x, M) = 0$. Cases with less than two included time-points, that is $|P_i(x, M)| < 2$, are left out. Note that in the

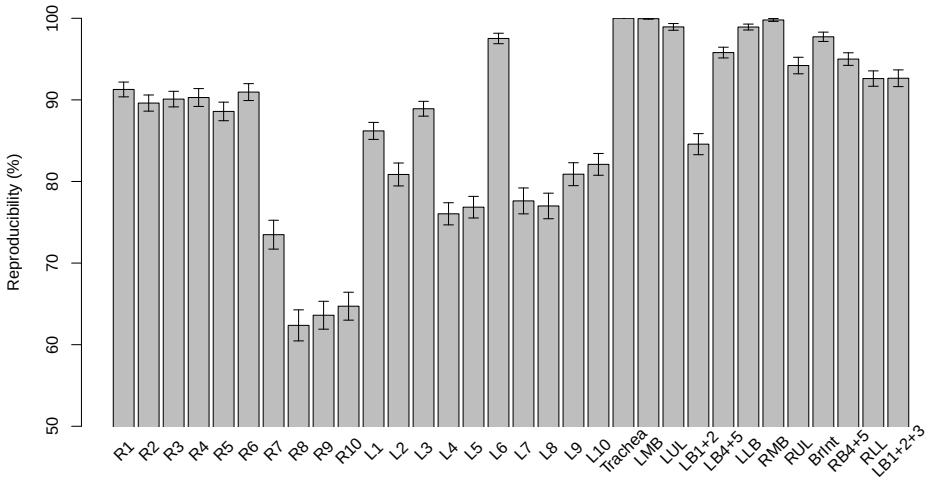
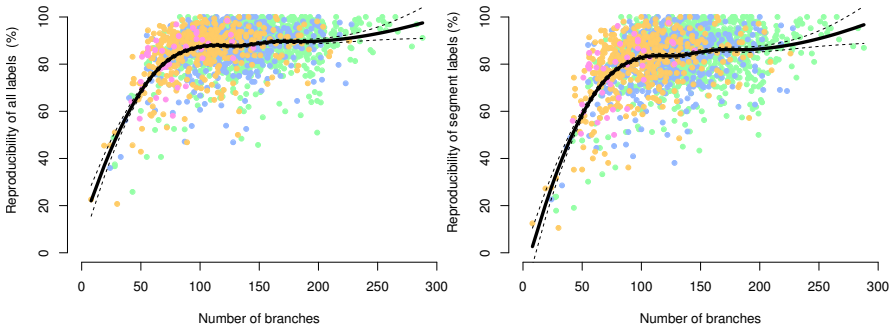


Figure 4.12: Reproducibility of the automatically assigned individual labels in the large longitudinal data set. Error bars indicate 95% confidence intervals.

case of two time-points, this definition of reproducibility is the same as the one defined in Section 4.3.4. Figure 4.12 shows a summary of the results.

On average, the reproducibility of the automatic labelling on the large longitudinal data set was $82.5 \pm 12.0\%$ on the segment labels, and $86.9 \pm 9.8\%$ overall.

There was a significant correlation between the reproducibility and severity of COPD ($\rho = -0.158, p < 0.001$ on all labels; $\rho = -0.163, p < 0.001$ on segment labels). Airway segmentations can be less complete in more diseased subjects (Pu et al. 2012; Diaz et al. 2010a), and we also observed a significant correlation between number of extracted branches and severity of COPD ($\rho = -0.444, p < 0.001$) and between number of extracted branches and reproducibility ($\rho = 0.287, p < 0.001$ on all labels; $\rho = 0.308, p < 0.001$ on segment labels). Figure 4.13 shows reproducibility plotted against number of segmented branches. To investigate whether the algorithm was truly sensitive to disease and not just missing branches, we generated a normalized reproducibility by subtracting the mean predicted reproducibility, predicted from the number of segmented branches using a locally weighted mean (Loess Curve) also shown in Figure 4.13, from the actual reproducibility. This normalized reproducibility did not significantly correlate with disease ($\rho = -0.043, p = 0.059$ on all labels; $\rho = -0.032, p = 0.166$ on segment labels).



(a) (b)

Figure 4.13: Reproducibility of all labels left and segment labels right as a function of the number of segmented branches. Colours indicate GOLD group, with GOLD 0 in green, GOLD 1 in blue, GOLD 2 in yellow, and GOLD 3 in pink. The black line is a locally weighted mean (Loess Curve) with 95% confidence intervals.

4.4 Discussion

We have presented a novel supervised algorithm for assigning anatomical branch labels in airway trees extracted from CT. Through detailed experimental validation we show that the performance of the algorithm is as good as the performance of the clinical experts. In particular, the accuracy and reproducibility of the algorithm is over 90% on the non-segment branches as well as on the L6 and R6 branches, and for many of the branches even over 95%. The labelling is fast and uses little memory, and easily runs on a standard laptop.

The labelling algorithm only uses centreline shape as input, which in many respects makes it robust to external factors such as disease. However, the hierarchical scheme of Figure 4.6 does make the labelling algorithm sensitive to missing branches and may cause difficulties with rare topologies. This could be handled by a more refined hierarchical labelling scheme, particularly one informed by an analysis of seen topologies, or of the airway sub-trees where the experts performed better.

We have chosen to use the geodesic tree-space distance between pairs of leaf-labelled trees. In principle, any other distance between leaf-labelled trees could have been used in its place, e.g. the weighted Robinson-Foulds metric (Robinson and Foulds 1979) which is a related distance measure in the same tree-space. The geodesic distance is attractive because, in addition to the fact that polynomial time algorithms are available, tree-space equipped with the geodesic distance allows for statistical tree-shape analysis of the airway trees (Feragen et al. 2013c). The strong performance of the labelling thus also works as a validation of the ability of the geodesic metric to represent

tree-shape differences in a way which is suitable for analysis of airway trees.

The labelling selection step of the labelling algorithm as shown in line 12 of Algorithm 3 selects the labelling of a new tree which minimizes the sum of distances to a training set of expert-labelled trees. This might introduce a bias of the labelling towards a "median tree", which could be problematic if labelled trees follow a multi-modal distribution within tree-space. Preliminary experiments, using the sum of distances to k nearest expert-labelled trees, did not result in significantly different results, which indicates this possible bias is of little consequence.

4.4.1 Robustness and applicability

The labelling algorithm attains statistically similar accuracy and higher reproducibility than the experts, and performs robustly in patients suffering from COPD. These qualities make the algorithm useful for clinical applications. One such application is analysis of airway dimensions measured from CT in studies of subjects with airway diseases such as COPD. One problem in performing such analysis is the variability introduced by including branches from different locations of the airway tree. Comparison of identically labelled airways in different subjects (Lederlin et al. 2012; Diaz et al. 2010a; Hasegawa et al. 2006) should decrease variation caused by measurement location and thus increase the ability of the measurement to capture signs of abnormalities. The labelling algorithm could also be applied to study the distribution of abnormalities within the lung in a group of patients, by lobe or segment. Comparison of measurements does not have to be limited to labelled branches, as branches in sub-trees of the labelled branches can also be included, e.g. through comparison of average measurements in generations relative to each specific label (Diaz et al. 2010a; Hasegawa et al. 2006). However, such an approach could be problematic if not all branches of each generation are found (Wielputz et al. 2013).

Our evaluation gives detailed insight into the difficulties of the labelling problem. It is noteworthy that the experts and the algorithm perform well in different parts of the airway tree. In particular, the algorithm is far more reproducible than the experts in the left upper and lower lobes (L1-L3 and L7-L10) ($[70\% - 87.5\%]$ versus $[57.1\% - 64.2\%]$). These branches are also the hardest to label according to expert accuracy ($< 60.1\%$). It is possible that biological variation of shape and topology confuses the experts, making their labels more random, which would lead to both low accuracy and reproducibility. The algorithm might either be more tolerant of biological variation or more consistent in the types of errors made. On the other hand, the experts perform better than the algorithm in the left middle and right lower lobes (L4-L5 and R7-R10). These branches belong to the sub-trees of LB4+5 and RLL, which are the least accurately found branches of the subset of branches used as steps in the hierarchical approach. Some amount of error is therefore

probably due to the hierarchical search strategy. It is possible that better results could be obtained by a more refined approach, for instance by searching for L4-L5 and R7-R10 label configurations within sub-trees depending on more than one choice of the LB4+5 and RLL branches.

It is interesting to compare the estimated reproducibility in the small and large data sets. In general the trends are largely the same. For instance in both data sets R6 and L6 are among the most reproducibly assigned segment labels and lower lobe segment labels are in general less reproducibly assigned than upper and middle lobe segment labels. The mean reproducibilities of the two data sets are also almost identical. It should be noted, however, that the automatic matching method (Petersen et al. 2011a) requires an accurate registration and unlike with the manual matching the overall topology and shape of the tree is not taken into account. Because of this the true reproducibility is probably underestimated in the larger data set. However, the larger data set also has relatively fewer severe COPD cases, which should mean the reproducibility is higher because of more completely segmented airway trees. Despite this, the similarity of the results indicate that the performance of the algorithm generalizes to new (albeit similar) data.

4.4.2 Labelling performance and COPD stage

Our experiments on the annotated dataset show that labelling reproducibility decreases significantly with increased COPD severity, while labelling accuracy does not. The difference in results may be caused by the mathematical definition of accuracy and reproducibility.

Labelling accuracy measures the ability of a method to agree with a human observer on the same segmented airway tree. If the underlying labelling "algorithm" used by the method and the human observer are identical, accuracy will be perfect and there will be no dependence on COPD severity even if both are physiologically incorrect. Reproducibility, on the other hand, measures the ability of *one* method to identically label two different segmentations of the same airway tree. If one of the segmentations is missing branches that are labelled by the other, then reproducibility cannot possibly be perfect. Thus, since the number of segmented branches depends on COPD level, it is expected that reproducibility depends on COPD level as well, while this is not necessarily true for accuracy.

This is supported by our reproducibility experiments on the large longitudinal data set. Here, reproducibility does again depend on COPD level. However, there is also a very clear correlation between reproducibility and the number of branches detected by the segmentation algorithm. When the effect of branch number is taken into account by subtracting the expected reproducibility based on number of branches, the correlation between reproducibility and COPD level disappears. We conclude that the correlation between COPD level and reproducibility is not an artefact of the labelling algorithm

directly, but a result of segmentation problems. This is very natural when many branches close to the true named branches are missing, and it has several consequences: if the named branch is missing, then any attempt to assign the corresponding label will fail, and if one but not both children of a named branch is missing, then the branch will appear longer in the segmentation than it should, making it harder to assign labels based on branch features such as shape.

Of course, one solution to the dependence on segmentation quality and, indirectly, disease, is to develop better segmentation algorithms. A more pragmatic approach, however, could be to introduce label probabilities based on geodesic airway tree distances, giving an option of assigning fewer labels when higher accuracy is needed, in a similar way as done in Lo et al. 2011. This could also be used to decrease the false positive rate on difficult branches.

4.4.3 Relation to alternative methods

We note that higher labelling accuracy percentages than ours are reported in the literature, 97.1%, 90%, 83% on all branch labels in Tschirren et al. 2005b, (high dose CT), Ginneken et al. 2008a, and Lo et al. 2011; and 77% on segment labels in Lo et al. 2011. There are several reasons for this. First, as noted above, we specifically aim to evaluate our performance on patients with disease, and our dataset consists of 77.5% COPD patients, while experiments found in the literature include much fewer, if any, subjects with an airway disease. A lower performance on our dataset is thus expected, since our experiments prove that labelling performance significantly depends on disease when segmentation problems are not accounted for.

Second, we aim to evaluate our ability to assign 20 segment labels. In comparison, both Lo et al. 2011 and Ginneken et al. 2008a use fewer than 20 segment labels (16 and 19, respectively) and more intermediate (easier) labels (13 as opposed to our 12), which presumably gives higher overall performance summaries. We note in particular that Lo et al. 2011 and Ginneken et al. 2008a leave out the segment label sets {L1-L2, L7-L8} and {L7}, respectively, which are also found challenging in our experiments.

Third, we wanted our evaluation to realistically reflect how well we can expect to perform on data which has never been labelled by an expert. To achieve this, we did not reject any assigned labellings, as opposed to Tschirren et al. 2005b and Lo et al. 2011, which aim to avoid performing uncertain labellings. In particular, the 97.1% success rate (Tschirren et al. 2005b) is among branches that have been labelled identically by three experts, which means that difficult labelling problems are weeded out of the experiment. Such an evaluation is only possible if the airway tree has already been labelled by three experts, and similar accuracies can naturally not be expected on unseen data, such as in clinical applications. In Lo et al. 2011, an estimate is made of

the probability of the label assignment, and here, a threshold can be applied to choose not to assign labels when certainty is low.

On average (including erroneous labellings), we assign 98.4% of 32 used labels, whereas Tschirren et al. 2005b, Ginneken et al. 2008a and Lo et al. 2011 assign only 71%, 93%, and 83% of the 29, 32, and 32 used labels, respectively. For segment labels specifically, we assign 94.9% of 20 used segment labels, whereas Lo et al. 2011 assigns 77% of the 16 segment labels used (our numbers are averaged over the 10 cross-validation runs). This variation in experimental setup makes it impossible to compare performance in a fair manner, because results on unassigned labels cannot be taken into account.

What we *can* conclude is that the proposed algorithm performs as well as or better than medical experts in terms of labelling accuracy and reproducibility. This is the best result we could possibly have hoped for given that our method is trained on labelling performed by medical experts. These conclusions are confirmed by our large-scale evaluation of reproducibility. Moreover, we quantify the dependence on performance on COPD level and show that any negative correlation between labelling performance and disease can be explained by segmentation error. To the best of our knowledge, no previous work has tested neither reproducibility nor dependence on disease, nor performed large-scale evaluations.

4.5 Conclusion

We present a new supervised algorithm for anatomical branch labelling of airway trees, based on geodesic tree-space distances between airway trees. Using the distances, the algorithm evaluates how well a suggested branch labelling fits with a training set of labelled airway trees, and chooses the optimal labelling. The labelling performance is robust in patients with COPD, and is comparable in performance to that of experts in pulmonary medicine and radiology. As the algorithm only uses branch centrelines and tree topology, we expect it to generalize to other data sets consisting of similarly complete segmentations of human adults. Its reproducibility and robustness in patients with COPD emphasizes its suitability for use in clinical studies of localized CT-based airway measurements.

Acknowledgements

This research was supported by the Lundbeck Foundation; AstraZeneca; The Danish Council for Strategic Research; The Danish Council for Independent Research | Technology and Production; Netherlands Organisation for Scientific Research (NWO); Centre for Stochastic Geometry and Advanced Bioimaging, funded by the Villum Foundation. Megan Owen acknowledges the support of the Fields Institute.

Chapter 5

Effect of inspiration in asymptomatic subjects

The work presented in this chapter is based on J. Petersen, M. M. W. Wille, L. L. Rakêt, A. Feragen, J. H. Pedersen, M. Nielsen, A. Dirksen, and M. de Bruijne (submitted). "Effect of inspiration on airway dimensions measured in maximal inspiration CT images". In: -.

Abstract

Objectives: Effect of inspiration on airway dimensions measured in voluntary inspiration breathhold scans is studied.

Methods: 961 subjects with normal spirometry were selected from the Danish Lung Cancer Screening Trial. Subjects were scanned annually for 5 years with low-dose CT. Automated software segmented lungs and airways, identified segmental bronchi, and matched airway branches in all scans of the same subject. Inspiration level was defined as segmented total lung volume (TLV) divided by predicted total lung capacity (pTLC). Mixed-effects models were used to predict relative change in lumen diameter (ALD) and wall thickness (AWT) in airways of generation 0 (trachea) to 7 and segmental bronchi (R1-R10 and L1-L10) from relative changes in inspiration level.

Results: Relative changes in ALD were related to relative changes in TLV/pTLC and this distensibility increased with generation ($p < 0.001$). Relative changes in AWT were inversely related to relative changes in TLV/pTLC in generation 3-7 ($p < 0.001$). Segmental bronchi were widely dispersed in terms of ALD (5.7 ± 0.7 mm), AWT (0.86 ± 0.07 mm), and distensibility ($23.5 \pm 7.7\%$).

Conclusions: Subjects who inspire deeper prior to scanning have larger ALD and smaller AWT. This effect is more pronounced in higher generation airways. Thus, inspiration level adjustment is needed to accurately assess airway dimensions.

5.1 Introduction

Airway dimensions measured from CT images are increasingly used to investigate remodelling of the airways caused by obstructive lung diseases (Hackx et al. 2012; Lederlin et al. 2012; Wielputz et al. 2013). Abnormalities, such as lumen narrowing and wall thickening, have been quantified using various measurements, examples include airway lumen diameter (ALD), airway wall thickness (AWT), airway wall area, and percentage wall area (Jong et al. 2005). Measures are different for different regions of the lungs and have therefore often been done in a few identified anatomical airway branches (Lederlin et al. 2012; Diaz et al. 2012; Hasegawa et al. 2006; Brown et al. 2001), for example the apical bronchus of the right upper-lobe, and the resulting conclusions have been extrapolated to hold for airways in general. However one particular segmental bronchus may not be representative of all and the pathology may not be homogeneously distributed throughout the airways (Hasegawa et al. 2006). Alternative methods which consider larger parts of the airway tree have been suggested. One such method is an indirect measurement of wall thickness at a virtual interior perimeter of 10 mm (Pi10). It is derived by assuming a linear relationship between the square root of airway wall area and interior perimeter (Nakano et al. 2000). Another such method is to average wall thickness measurements over all detectable airways with a specific ALD (Dijkstra et al. 2013; Achenbach et al. 2008). Both these approaches are influenced by changes in the wall and the lumen, making them difficult to interpret. An alternative is to average over all detected airway branches in a specific airway generation or airway generation range (Wielputz et al. 2013; Petersen et al. 2011b). Airway generations can be defined from the trachea by counting bifurcations in the airway tree; the idea being that higher generation corresponds to more peripheral and thus typically smaller airways.

In order to accurately assess airway dimensions, and use the results in a clinical setting, it is necessary to know the influence of the inspiratory level (Brown et al. 2001). It is known that airways are distensible (Diaz et al. 2012; Brown et al. 2001; Matsuoka et al. 2008; Brown et al. 1994; Scichilone et al. 2000; Scichilone et al. 2001; Baldi et al. 2010), that is, they follow the expansion and contraction of the lungs during the breathing cycle. It has not been established yet, however, by how much the airways distend, whether this distensibility depends on location in the lung, or what it means for variability of measurements in breath-hold scans. This study evaluates the effect of inspiration level differences by airway generation and anatomical segments on a large data set including multiple longitudinal scans of subjects without COPD.

5.2 Material and methods

5.2.1 Study population

The Danish Lung Cancer Screening Trial (DLCST) (Pedersen et al. 2009) is a 5-year study investigating the effect of CT screening on lung cancer mortality. It consists of 4,104 participants, who were randomly selected for either CT screening ($n = 2,052$) or no screening. Participants had to be 50-70 years of age, current or ex-smokers with a minimum of 20 pack-years. The DLCST was approved by the Ethics Committee of Copenhagen County and fully funded by the Danish Ministry of Interior and Health. Approval of data management in the trial was obtained from the Danish Data Protection Agency. The trial is registered in ClinicalTrials.gov's Protocol Registration System (identification no. NCT00496977). All participants provided written informed consent.

The present post hoc analysis included the subset of the 2,052 subjects in the CT screening arm, with at least three analysable scans and without airflow limitation, due to evidence of reduced airway distensibility in such populations (Diaz et al. 2012).

5.2.2 Spirometry

Spirometry was performed annually during a period of 5 years, according to recommendations by the European Respiratory Society (Miller et al. 2005) using a computerized system (Spirotrac IV; Vitalograph); however, no bronchodilation was applied. Measurements of forced expired volume in 1st second (FEV1) and forced vital capacity (FVC) were averaged over all annual time points and used to define airflow limitation as having an FEV1 to FVC ratio of less than 0.7. FEV1 as a percentage of the predicted normal value (FEV1 % predicted) was computed according to European reference equations (Pellegrino et al. 2005).

5.2.3 CT scans

The screening group was CT scanned annually during a period of 5 years, using a multi-slice CT scanner (16 rows Philips Mx 8000, Philips Medical Systems). Scans were performed in supine position after full inspiration with caudo-cranial scan direction including the entire ribcage and upper abdomen (field of view was 40 cm) with a low dose (120 kV and 40 mAs). Spiral data acquisition with the following parameters was used: Section collimation 16 x 0.75 mm, pitch 1.5, rotation time 0.5 second. Participants were instructed to first hyperventilate three times and hereafter inhale maximally and hold their breath during scanning. Each image was reconstructed with two kernels: thick (3 mm) and thin (1 mm) slice thickness using soft and hard algorithms (kernel C and D), respectively. The experiments of this study used the thin slice thickness unless otherwise is specified.

5.2.4 Lung extraction and measurements

The lungs were extracted from the images using a fully automatic region growing approach, using an upper threshold of -400 Hounsfield units (Lo 2010). The results have been validated by two medical experts (Ashraf et al. 2011) and images were discarded as non-analysable in case the lungs were not fully contained in the image or if bowels were included in the segmentation.

Percentage of low attenuation area less than -910 and -950 Hounsfield units (%LAA-910 and %LAA-950), and the Total Lung Volume (TLV) were computed from the segmented volumes. The low attenuation areas were computed based on the images with thick slice thickness. Predicted normal Total Lung Capacity (pTLC) was computed according to Quanjer et al. 1993. TLV and pTLC have been used to quantify inspiration level as TLV/pTLC.

5.2.5 Airway extraction and measurements

The airway trees were initially extracted using a locally optimal path approach (Lo et al. 2009) and then input to a graph based method (Chapter 2), which outputs triangulated meshes of the interior and exterior wall surfaces. The approach has been validated on DLCST data showing relatively complete airway trees with very few false positives (Lo et al. 2009) and sub-voxel accuracy when compared to manual annotations and the COPD gene phantom (Chapter 2).

The airway centreline tree was extracted from the interior wall surface using the front propagation method described in Lo et al. 2012.

The anatomical names of the segmental bronchi were assigned based on the geometric tree space approach of Chapter 4, which has been shown to be statistically similar in accuracy and reproducibility to medical experts on DLCST data.

Deformable image registration (Gorbunova et al. 2012) was used to match individual airway branches in multiple images of the same subject. Only branches matched at all time points were included. The details of the matching approach are given in Chapter 3.

Lumen diameter and wall thickness were computed at each of the mesh surface points, from the distance to the nearest point on the centreline. Finally, these measures of ALD and AWT were averaged in each branch.

5.2.6 Statistical methods

The influence of the explanatory variables TLV/pTLC, age, and branch position on the outcome variables ALD and AWT were analysed in four multiple regression models. Two models were used for each outcome, investigating position effects either in the form of generation (0 to 7) or segmental bronchus label (R1-R10 or L1-L10). TLV/pTLC, ALD, and AWT were log transformed to model relative relationships. Age groups, in the form of quartiles of the

Characteristic	Mean ± SD or No. (%)
Age, years	59.2 ± 4.6
Male gender	534 (55.6)
BMI, kg/m2	26.1 ± 3.9
Smoking history, pack-years	34.5 ± 12.2
Currently smoking	599 (62.3)
FEV1 % predicted	94.5 ± 12.6
FEV1/FVC ratio	0.75 ± 0.03
TLV/pTLC	0.91 ± 0.12
%LAA-910, %	12.4 ± 10.0
%LAA-950, %	0.74 ± 0.89

Table 5.1: Demographic, clinical and CT scan characteristics

mean age over visits, were included as categorical variables. We did not include subject sex and height as potential confounders, as the effects of these are already removed from the lung volume through the normalization by pTLC. Subject specific and (intra-subject) branch specific differences in airway dimensions were modelled as random effects (slope and intercept with respect to log(TLV/pTLC)). All possible interactions of the explanatory variables were tested. The Appendix contains a more detailed description of the statistical analysis.

5.3 Results

Of the 2,052 subjects in the screening arm of the DLCST, 53 left prior to the first lung function test, and 983 (49%) of the remaining subjects were excluded due to airflow limitation. 55 subjects had less than three analysable scans, giving a total of 63, 220, and 678 subjects with respectively 3, 4, and 5 annual scans. Table 5.1 contains demographic details of the study population. 204 out of the 4692 scans of subjects without airflow limitation were deemed non-analysable due to technical difficulties, such as an incomplete scan of both lungs and/or inclusion of bowel air into the lung segmentation.

The variation in inspiration level was considerable with a standard deviation of TLV/pTLC within each subject (mean ± standard deviation) of: 5.4±3.2%. The standard deviation of the subject specific means of TLV/pTLC was 13.3%.

The analysis included a total of 106,162 airways, measured on average 4.6 times. Figure 5.1 shows the mean number of airways found and matched in each subject by generation.

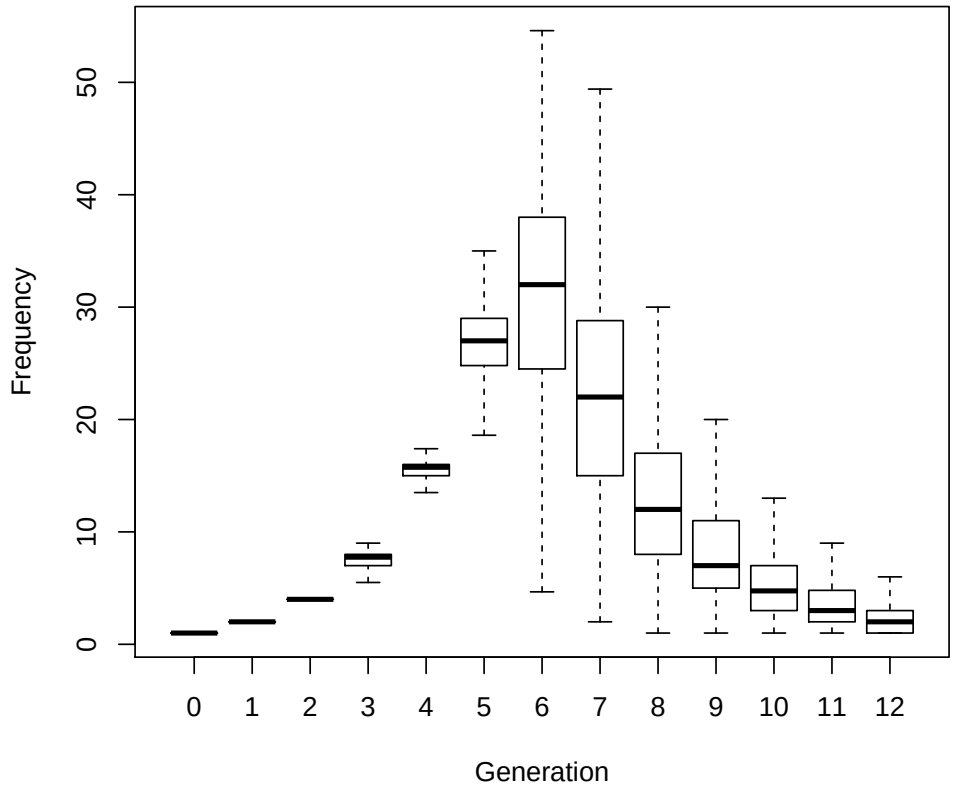


Figure 5.1: Number of analysed airway branches by generation. Line at median, box at first and third quartile and whiskers at box plus 1.5 times the interquartile range.

5.3.1 Statistical analysis

Interactions between age and log(TLV/pTLC) and between age and position (generation or segmental bronchi) did not improve the predictions of either log(ALD) or log(AWT) ($p > 0.084$) and were removed from the models. Position, log(TLV/pTLC) and the interaction of position and log(TLV/pTLC) were all significant predictors of both log(ALD) and log(AWT) ($p < 0.001$). Age was a significant predictor of log(ALD) ($p < 0.001$), but only borderline significant for log(AWT) ($p = 0.057$).

ALD increased with age; ALD in airways of the same generation were 1.2% ($p < 0.05$), 0.8% ($p = 0.177$), and 2.3% ($p < 0.001$) larger in the older quartiles (55-59, 59-63, and 63-73 years respectively) compared to the youngest quartile (51 to 55 years). The same trend, albeit smaller, was observed for AWT, with corresponding percentages of 0.3% ($p = 0.147$), 0.4% ($p = 0.105$), and 0.4% ($p = 0.057$).

Figure 5.2 shows ALD plotted against AWT in every examined posi-

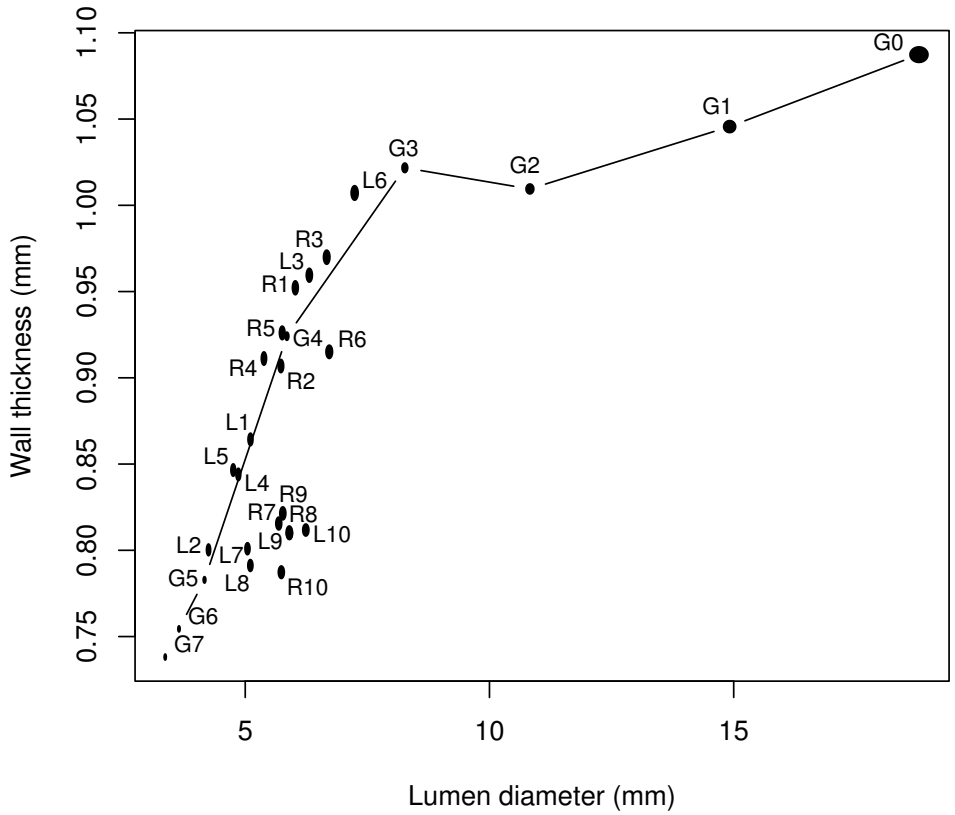


Figure 5.2: Lumen diameter plotted against wall thickness for the segmental bronchi (R1-R10, L1-L10) and generation 0 to 7 (G0-G7). The ellipses indicate standard errors in both dimensions.

tion (generation 0-7, R1-R10, and L1-L10), as predicted by the models for subjects of the youngest age quartile (51 to 55 years) and at ideal inspiration (TLV=pTLC). AWT and ALD decreased with increasing generation ($p < 0.001$). The segmental bronchi were widely dispersed in terms of generation number with mean $\pm$ standard deviation of 4.7 ± 1.0 ranging from 3.0 ± 0.1 for L6 to 7.0 ± 0.4 for R9, in terms of lumen diameter with mean $\pm$ standard deviation of 5.7 ± 0.7 mm ranging from 4.23 [4.20-4.27] mm for L2 to 7.22 [7.16-7.28] mm for L6, and in terms of wall thickness with mean $\pm$ standard deviation of 0.86 ± 0.07 mm ranging from 0.79 [0.78-0.79] mm for L8 to 1.00 [1.00-1.01] mm for L6.

ALD increased with inspiration level ($p < 0.001$) and increasingly so with generation ($p < 0.001$). We define lumen distensibility as the degree to which ALD increases with an increase in inspiration level. For instance a 100% volume increase in completely yielding airways should correspond to about 26% distension in each direction, assuming equal distension in all directions, thus leading to a 26% increase in ALD. As seen in Figure 5.3, this was observed

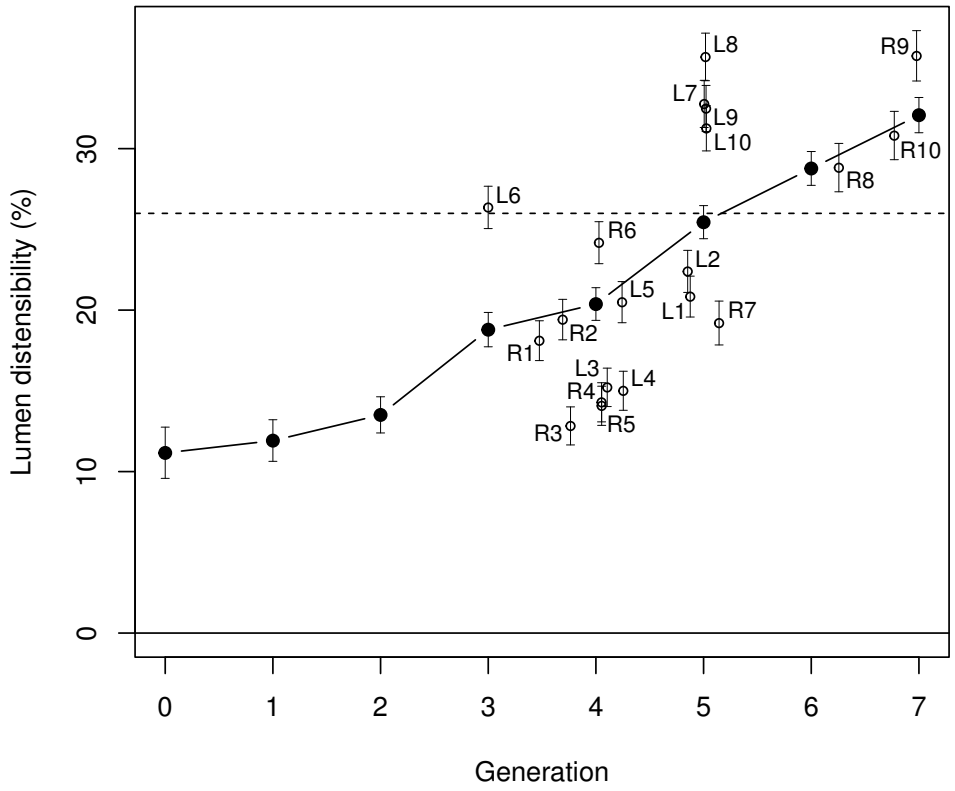


Figure 5.3: Lumen distensibility at a 100 % increase in inspiration level by generation (connected and filled circles) and segmental bronchi (R1-R10 and L1-L10). The dotted line indicates a distensibility equal to lung parenchyma, and error bars indicate standard errors.

in approximately the airways of generation 5 and beyond. The mean ± standard deviation of the distensibility of the segmental bronchi was $23.5 \pm 7.7\%$. The least distensible was R3 with a change of 12.8 [11.6-14.0] %, whereas R9 changed 35.7 [34.2-37.3] %. Larger size does not always indicate lower distensibility, even though diameter decreases and distensibility increases with generation, as evidenced by L6 having the largest estimated diameter and thickest walls, but also a relatively high distensibility. Distensibility was greater in the segmental bronchi of the lower lobes (L6-L10 and R6-R10), where the mean was 29.7% with estimates ranging from 19.2% to 35.7% as compared to the segmental bronchi of the lingular, middle and upper lobes (L1-L5 and R1-R5), where the mean was 17.3% with estimates ranging from 12.8% to 22.4%.

AWT, on the other hand, decreased with inspiration level (with the exception of AWT in the airways of generation 0-2) ($p < 0.001$). We define wall thinning, similarly to distensibility but in the opposite direction, as the degree to which AWT decreases with an increase in inspiration level. Figure 5.4

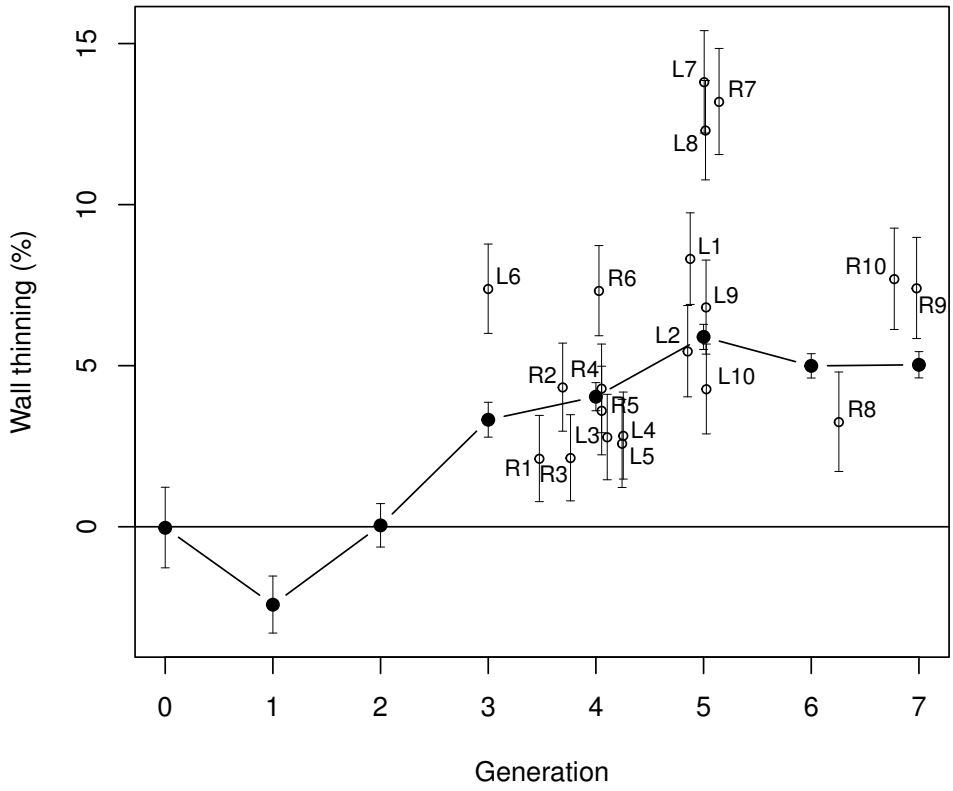


Figure 5.4: Wall thinning at a 100 % increase in inspiration level by airway generation (connected and filled circles) and segmental bronchi (L1 to L10 and R1 to R10). Error bars indicate standard errors.

shows that wall thinning increases with generation, but flattens out at around 5%, suggesting that a doubling of inspiration level will lead to a decrease in AWT of 5% in the more peripheral airways. The mean ± standard deviation wall thinning of the segmental bronchi was $6.1 \pm 3.6\%$. The least thinning was R1 with a change of 2.1 [0.8-3.5] %, whereas L7 changed 13.8 [12.2-15.4] %.

The subject specific and intra-subject branch specific variation in ALD and AWT as estimated by the random effects and the model residual standard deviation are for completeness listed in the Appendix.

5.4 Discussion

The difference between typical abnormal and normal airway measurements is small, and the groups are largely overlapping (Lederlin et al. 2012; Diaz et al. 2012; Achenbach et al. 2008). Our results show that variation in level of inspiration, it being either inter- or intra-individual, may alter the diagnosis

from healthy to sick or vice versa. For instance in COPD larger values of percentage wall area and Pi10 are usually associated with a higher degree of airflow limitation. However, similar measures may be found in a subject who inspired less during scanning; a decrease in inspiration level results in a smaller lumen and thicker wall, which also translates into a higher percentage of wall area. Although the situation is slightly more intricate for Pi10, a reduction in inspiration level will also make the Pi10 measurement larger. Moreover, as distensibility changes with airway generation and anatomical segment, the change in Pi10 will depend on which airway branches are included in the regression. A similar conclusion can be drawn regarding the effect of inspiration level differences on the results of approaches that average measurements over all detected airways with a certain interior diameter (Dijkstra et al. 2013; Achenbach et al. 2008).

Our results indicate that the effect of changes in inspiration level on measured airway properties is highly dependent on the position of the branch in the airway tree. Conclusions based on measurements of a limited number of branches (Lederlin et al. 2012; Diaz et al. 2012; Hasegawa et al. 2006; Brown et al. 2001) may therefore not be representative of the airway as a whole. More representative measurements may be achieved by combining complete sections of the airway tree by generation and by standardizing for the level of inspiration.

Adjustment for the effects of inspiration on airway dimensions can be done using a model such as ours, if multiple scans are available on each subject. It is, however, difficult to extract general guidelines on how to adjust if this is not the case. An important finding in this respect is that the average airway at generation 5 and beyond, or the average segmental bronchi, can be assumed to be completely distensible. In other words these airways follow the expansion and contraction of the lung, offering no more resistance than the lung parenchyma. We therefore suggest that ALD measurements in these airways are standardized by dividing by the cubic root of TLV/pTLC. AWT in the same airways in our study only becomes about 5 % thinner at a 100 % change in inspiration level, which is far less than what would be expected if the wall volume was constant. This might be because the changes are below the resolution of the system, or partial volume effects could cause the thinning to appear less than it actually is. We thus refrain from recommending an adjustment of AWT until future studies have established the exact nature of this relationship.

The study includes a large, but selected population of current and former heavy smokers in the age range from 50 to 70 years, with a smoking history of at least 20 pack-years, and with normal lung function. Thus, the conclusions cannot without reservations be expanded to include subjects that are younger, never-smokers or suffering from lung disease.

A low-dose scanning protocol was used, which may have reduced measurement accuracy (Hackx et al. 2012). Several studies have examined or performed

airway analysis on low dose scans (Dijkstra et al. 2013; Tschirren et al. 2005a; Lutey et al. 2013; Jong et al. 2006), and concluded that this has reasonable accuracy. Measurement uncertainty, such as that caused by the use of low dose scans, is included in our statistical models as the residual error terms (Appendix). Moreover the methods used in our study have all been validated with the DLCST data previously showing that a large part of the airway tree can be automatically and repeatedly found and measured accurately and reproducibly using low-dose techniques (Lo et al. 2009 and Chapter 2,3, and 4).

The number of found and matched airway branches roughly doubles with generation, as expected, until generation 6. This suggests that at this point the airways begin to fall below the resolution limit imposed by the system and branches are missed. Thus, our sampling is likely biased towards slightly larger airways and results may not accurately reflect the average properties of generation 6 and beyond. The resulting bias in distensibility and wall thinning is less clear, and the results of generation 6 and 7 should be interpreted with care.

In conclusion, the results show that subjects who inspire deeper prior to scanning tend to have larger lumen diameter and thinner walls and more pronouncedly so in more peripheral airways. This effect makes it necessary to adjust for inspiration level, to accurately assess airway dimensions as indicators of obstructive lung diseases.

Acknowledgements

The authors wish to thank Pechin Lo and Vladlena Gorbunova for developing initial versions of the software used for airway extraction and matching and AstraZeneca Sweden and the Netherlands Organisation for Scientific Research (NWO) for financial support.

Appendix

Statistical methods supplement

The statistical models take the form:

$$Y_{i,j,k} = \sum_{g=1}^{p} \left(\beta_{1,g} \times G_{g,i,j} + \beta_{2,g} \times G_{g,i,j} \times \log\left(\frac{\mathrm{TLV}_{i,k}}{\mathrm{pTLC}_i}\right)\right) + \sum_{a=1}^{4} (\beta_{3,a} \times A_{a,i}) + \alpha_i + \alpha_{i,j} + (\gamma_i + \gamma_{i,j}) \times \log\left(\frac{\mathrm{TLV}_{i,k}}{\mathrm{pTLC}_i}\right) + \epsilon_{i,j,k},$$

where the index i refers to subject, j to the branch within the subject, k to the visit number and g to the location of the branch. The outcome variable Y is either log(ALD) or log(AWT). $\beta_{1,g}$, $\beta_{2,g}$ and $\beta_{3,a}$ for $g \in \{1,2,3,4\}$ are

the coefficients of interest, describing the outcome variable's relationship with location of the branch, interaction of location and inspiration level, and age of the subject. The random intercepts α_i and $\alpha_{i,j}$ and slopes γ_i and $\gamma_{i,j}$ are assumed to be independent, normally distributed with zero mean across subjects and within subjects respectively. The residuals $\epsilon_{i,j,k}$ follow independent zero-mean normal distributions. The categorical variables $G_{g,i,j} \in \{0, 1\}$ describe whether a branch is a segmental bronchus, that is R1-R10 or L1-L10 ($P = 20$) or whether it belongs to generation 0 to 7 ($P = 8$). $A_{a,i} \in \{0, 1\}$ Are the four categorical variables indicating age quartile of the subject. The models are fitted using restricted maximum likelihood via the lme4 package (Bates et al. 2012) for the R programming language (R Core Team 2012).

Statistical results supplement

The standard deviation of the mean ALD in a generation and ALD in a segmental bronchus were 9.8% and 11.0% between subjects belonging to the same age group and with the same inspiration level. For AWT the corresponding numbers were 3.4% and 4.3%. The intra-subject standard deviation of ALD at different time points, standardized for inspiration level, was 27.7% in a given generation and 17.4% in a given segmental bronchus. For AWT the corresponding numbers were 8.5% and 8.5%.

The standard deviation of the mean lumen distensibility of airways in the same generation and for a given segmental bronchus was respectively 26.5% and 17.4% between subjects of the same age group. In terms of wall thinning the corresponding numbers were 6.7% and 5.1%. The intra-subject standard deviation of lumen distensibility in each generation and each segmental bronchus were 10.5% and 6.7% respectively. For wall thinning the numbers were 6.8% and 9.9%.

The standard deviations of the residuals were 6.6% and 4.8% in terms of ALD with generation and segmental bronchus respectively describing position. For AWT the numbers were 10.8% and 10.9%.

Chapter 6

Distensibility, lumen diameter, and wall thickness in subjects with chronic obstructive pulmonary disease

The work presented in this chapter is based on M. M. W. Wille, J. Petersen, S. B. Shaker, A. Dirksen, L. T. Skovgaard, J. H. Pedersen, and M. de Bruijne (to be submitted). "Airway distensibility, lumen diameter, and wall thickness in COPD - based on CT airway segmentation in the Danish Lung Cancer Screening Trial". In: -.

Abstract

Background: The aim of this study was to investigate the influence of chronic obstructive pulmonary disease (COPD) on the dynamic change in airway calibre and the airway wall thickness, using consecutive inspiratory CT scans from the Danish Lung Cancer Screening Trial (DLCST).

Materials and methods: From the screening arm of DLCST 1,869 participants with at least three of five planned annual CT scans were included, all of whom underwent lung function testing annually. Based on spirometry and according to the Global Initiative for COPD (GOLD), participants were divided into groups 0-4 of increasing airflow limitation (AFL) as an indicator of disease severity in COPD. Repeat scans were registered, airways were segmented and labelled as generation 0-7, and finally airway lumen diameter and wall thickness were measured. Airway distensibility was defined as the relative change in lumen diameter divided by the relative change in total lung volume. Groups were compared using two mixed-effects models with the following specifications: Outcome = log(airway lumen diameter) or log(airway wall thickness) and explanatory variables = airway generations 0-7, log(total lung volume (TLV)), participant age quartiles, and GOLD groups 0-4.

Results: Lumen diameter decreased with generation and disease severity ($p < 0.001$). Distensibility increased with generation and decreased with disease severity ($p < 0.001$). Wall thickness decreased with generation and increased with disease severity ($p < 0.001$).

Conclusion: COPD is affecting large airway distensibility, lumen diameter and wall thickness. Changes are more pronounced toward smaller airways, which suggests comparable changes are present in small airways with lumen diameters below 2 mm.

6.1 Introduction

Chronic obstructive pulmonary disease (COPD) is one of the leading causes of death and disability (Lopez et al. 2006), particularly among heavy smokers. COPD consists of chronic bronchitis and emphysema, which may both lead to airflow limitation.

The pathophysiologic background of COPD has traditionally been divided into two distinct pathways: 1) emphysema, characterized by parenchymal destruction causing loss of elastic recoil, and 2) chronic bronchitis, in which inflammation and fibrosis of the bronchioles give rise to narrowing of the small airways (Ciba 1959). The combined effect of these two distinct pathways is the explanatory background for the clinical presentation of COPD; decreased maximal expiratory airflow, hyperinflation and gas trapping due to increased airway resistance and premature airway closure (Cosio et al. 1978; Cosio and Guerassimov 1999; Hogg et al. 1968). It has been argued that the process that causes airway limitation is inflammatory above all, causing thickening of all layers of the bronchial wall and narrowing of the airway lumen (Hogg et al. 1968; Hogg et al. 1969) as well as accumulation of mucus (Hogg et al. 2004).

Emphysema is defined as an abnormal and permanent enlargement of the air spaces distal to the terminal bronchioles and destruction of alveolar and bronchial walls (Ciba 1959), and even though emphysema is a diagnosis based upon pathology, it may be assessed by quantitative computed tomography (CT) measuring density of the lung. This technique has been validated against pathology (Gould et al. 1988) and has been used in clinical trials of new treatments in alpha1-antitrypsin deficiency (Dirksen et al. 1999; Stolk et al. 2012)

Large airways are anatomically defined as airways containing cartilage and glands in the wall and are located up to approximately 10th generation. The airways located within generation 0-7 of the airway tree have an average diameter of 2 mm or more (Weibel 2009). The spatial resolution of low-dose CT sets a limit to airways eligible for analysis at approximately the 7th generation, which therefore prevents this study from investigating small airways.

Emphysema leads to airflow limitation due to dynamic collapse of airways during expiration caused by decreased lung recoil and loss of alveolar attachments that keep airways open during expiration in the normal lung (Diaz et al. 2012; Mead et al. 1967).

Spirometry gives an indication of the effect of emphysema and chronic bronchitis on lung function, but cannot distinguish between the two components of COPD. Several studies have argued that the two pathophysiologic pathways independently contribute to the clinical appearance (Grydeland et al. 2010; Nakano et al. 2000) resulting in distinct phenotypes of COPD formerly referred to as blue bloaters (obstructive bronchiolitis predominance) and pink puffers (emphysema predominance), respectively (Rutten et al. 2011).

Several factors influence CT airway measurements, and among the more

important is level of inspiration (Jong et al. 2005). Even if instructions to participants are standardized when CT scans are performed, the level of inspiration may vary considerably between scans (Gierada et al. 2001). In a previous Chapter 5 we showed that airway lumen diameter (ALD) increases and wall thickness (AWT) decreases with inspiration level and that this distensibility effect is more pronounced in higher generation airways.

The aim of this study was to investigate the combined influence of COPD and inspiration level on dimensions (ALD and AWT) of airways in different generations taking subject age into account. The inevitable variation in inspiration level between repeat scans of the same subject will be used to establish the effect of COPD on dynamic changes of airways close to maximum inspiration.

6.2 Material and methods

6.2.1 Study population

Participants who were randomized to annual low-dose CT in the Danish Lung Cancer Screening Trial (DLCST) comprise the study population. DLCST is a 5-year trial investigating the effect of screening on lung cancer mortality (Pedersen et al. 2009). Participants were recruited by newspaper advertisements. A total of 4,104 participants were enrolled and randomised to either annual screening with low-dose CT (n=2,052) or a control group (n=2,052), in which no CT screening was offered. Participants were men and women, 50-70 years of age, without lung cancer related symptoms, all current or ex-smokers who had smoked at least 20 pack-years, with forced expiratory volume in one second (FEV1) of at least 30 % of predicted normal at baseline. Ex-smokers had to have quit after the age of 50 years and less than 10 years before inclusion. The database included no information on use of bronchodilators and steroids.

At annual screening rounds, smoking habits were recorded, carbon monoxide level in exhaled breath was measured, and spirometry was performed. The trial was performed in one institution: Gentofte Hospital in Copenhagen, Denmark.

The DLCST was approved by the Ethics Committee of Copenhagen County and funded in full by the Danish Ministry of Interior and Health. Approval of data management in the trial was obtained from the Danish Data Protection Agency. The trial is registered in Clinical Trials.gov Protocol Registration System (identification no. NCT00496977). All participants provided written informed consent.

6.2.2 Imaging

All CT scans were performed on the same multi-detector CT scanner with 16 detector rows (Philips Mx 8000, Philips Medical Systems, Eindhoven, The

Netherlands). Scans were performed in supine position at full inspiration with a caudo-cranial direction and included thorax and upper abdomen. Participants were instructed by voice recordings to first hyperventilate three times, and then take a deep breath and hold it during scanning (10-12 sec.). A low-dose technique was applied (120 kV and 40 mAs) with the following acquisition parameters: field of view 40 cm; collimation 16 x 0.75 mm; rotation time 0.5 second; and pitch 1.5. Each image was reconstructed with both 3 and 1 mm slice thickness using soft and hard reconstruction algorithms (kernel C and D) respectively. The experiments used the 1 mm slice thickness images unless otherwise specified. The scanner was calibrated daily for air and with regular intervals for water according to the recommendations of the manufacturer.

6.2.3 Image analysis

All CT scans were analysed by the Image Group at the Department of Computer Science, University of Copenhagen, using in-house developed software designed to segment the lung and airways. The lung segmentation algorithm (Lo 2010) starts by detecting the trachea in the top part of the image. A special region growing algorithm segments the trachea and right and left main bronchi, which are not included in the total lung volume. Subsequently, the lungs are segmented starting from the end of the main bronchi using a competing region growing algorithm and a threshold of -400 HU for the interface between lung and extra-pulmonary tissue. The analysis is fully automated, and afterwards segmentations were inspected visually, and scans were discarded if the lungs were not fully included, or if bowel air was included in the segmentation.

Lung density was characterized from the images with 3 mm slice thickness by two low attenuation area (LAA) parameters: %LAA-910 was defined as the percentage of lung pixels with density $<$ -910 HU and %LAA-950 was defined as the percentage of lung pixels with density $<$ -950 HU.

Predicted normal Total Lung Capacity (pTLC) was computed according to Quanjer et al. 1993. TLV and pTLC was used to quantify inspiration level as TLV/pTLC (Dirksen 2008; Shaker et al. 2012).

The extraction of the airway wall surfaces was a two-step process. First, the airway centreline was extracted from the automatically detected seed point in the trachea by iteratively adding local optimal paths that most resemble airway centreline based on a statistical model derived from a training set. The process was described in Lo et al. 2009, and it was validated on a random subset of 10 scans from DLCST by an experienced observer. Out of an average tree of length 258 cm (excluding the trachea and two main bronchi), only 3.8% of branches were judged to contain errors. Secondly, the centreline was used as input in a graph based approach, which simultaneously detected both the interior and the exterior wall surfaces using image gradients. The method

was described and validated in Chapter 2, showing sub-voxel accuracy when compared to manual annotations.

Deformable image registration was used to match each airway branch in repeat scans of the same subject. We used a mass preserving lung registration method (Gorbunova et al. 2012) to register all images of the same subject to the one most central in time. Only branches that were found in all scans were included. The details of the approach were given in Chapter 3.

6.2.4 Lung function testing

During the 5 years of the study spirometry was performed annually, according to recommendations by the European Respiratory Society (Miller et al. 2005) using a computerized system (Spirotrac IV; Vitalograph, Buckingham, UK). No bronchodilation was applied. Measurements of Forced Expiratory Volume in 1 second (FEV1) and forced vital capacity (FVC) were expressed in absolute values and for FEV1 as percentage of predicted normal (pFEV1) according to European reference equations (Pellegrino et al. 2005). These values were then averaged over all visits, and airflow limitation (AFL) was defined as FEV1/FVC < 0.7. According to the Global Initiative for Chronic Obstructive Pulmonary Disease (GOLD) (GOLD 2013) each participant was classified in five stages of increasing AFL: No AFL (FEV1/FVC < 0.7), GOLD 1 (pFEV1 $> 80\%$), GOLD 2 ($50\% <$ pFEV1 $<80\%$), GOLD 3 ($30\% <$ pFEV1 $< 50\%$) and GOLD 4 (pFEV1 $< 30\%$). Due to small numbers we decided to combine GOLD 3 and 4 in one GOLD 3+4.

6.2.5 Statistics

Influence of explanatory variables TLV/pTLC, disease severity (No AFL and GOLD 1, 2, and 3+4), age and measurement position (generation 0 to 7) on airway outcome variables ALD and AWT were analysed in two mixed effects regression models. TLV/pTLC, ALD, and AWT were log transformed to model relative relationships. Age groups, in the form of quartiles of the mean age of visits, were included as a categorical variable to investigate ageing effects. Subject and (intra-subject) branch specific differences in airway dimensions were modelled as random effects (slope and intercept with respect to log(TLV/pTLC)). All possible interactions of the explanatory variables were tested. The appendix contains a more detailed description of the statistical models.

Airway distensibility, defined as the percentage ALD increases relative to a percentage increase in inspiration level, can be computed directly from the estimated coefficients of log(TLV/pTLC).

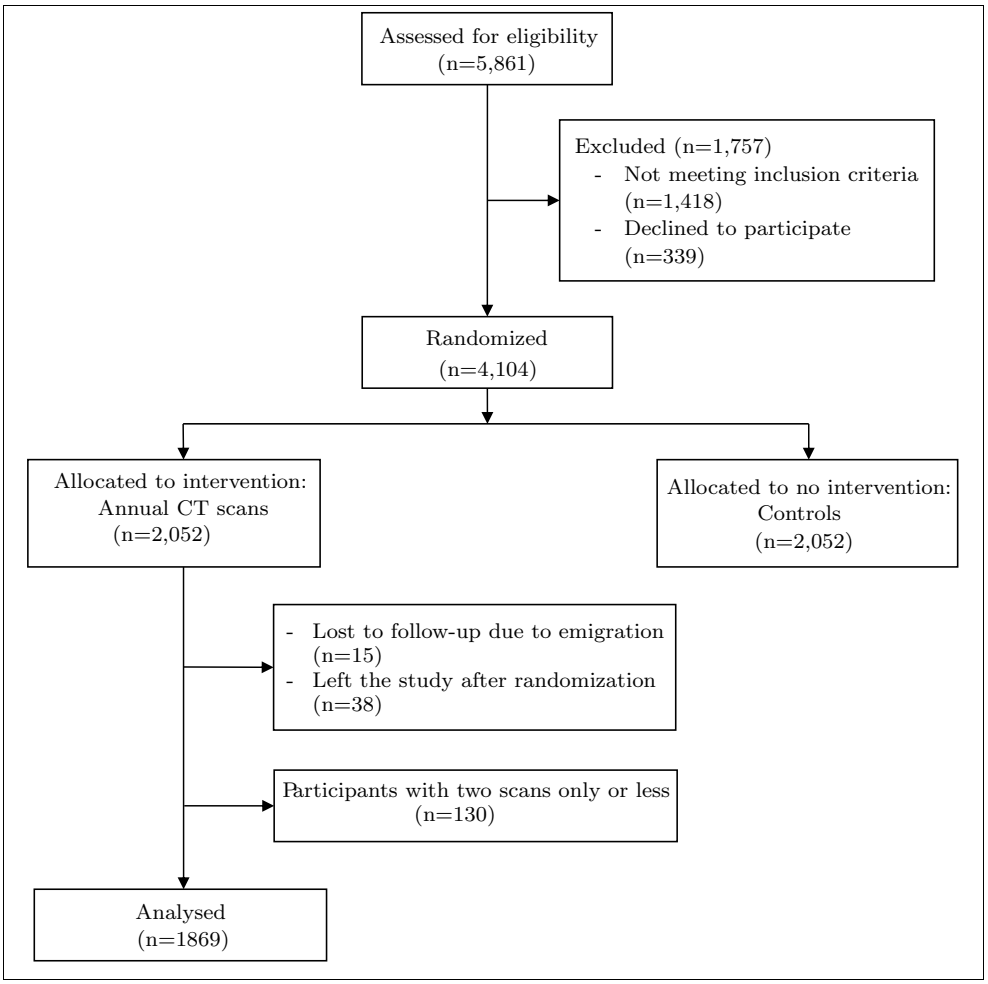


Figure 6.1: Flowchart of participating subjects. * = Subjects either left the study prematurely or there were technical difficulties with the scans, such as an incomplete scan of the lungs, or bowel-air being included in the segmentation.

6.3 Results

6.3.1 Description of the study population

CT scans from 1869 participants were included (Figure 6.1). Three, four, and five scans were included from 119, 444, and 1306 subjects, respectively. Most (52.2%) had no airflow limitation (No AFL), and only 2.1% had severe AFL (GOLD 3+4). There were slightly fewer women than men.

Table 6.1 shows characteristics of the population as mean values across all visits. Participants with AFL were older, had lower body mass index (BMI), had smoked more, had larger lungs as measured by the TLV/pTLC ratio, and had lower pFEV1 than subjects with no AFL.

Characteristics	No AFL (n = 961)	GOLD 1 (n = 488)	GOLD 2 (n = 382)	GOLD 3+4 (n = 38)	p-value
Age, years	59.16 ± 4.58	60.39 ± 4.99	60.76 ± 4.60	63.24 ± 4.66	$< .001$
Male sex	534(55.6)	279(57.2)	217(56.8)	24(63.2)	0.776
BMI, kg/m^2	26.07 ± 3.88	24.64 ± 3.39	25.08 ± 3.86	25.05 ± 4.88	$< .001$
Pack years	34.53 ± 12.19	36.63 ± 14.16	38.99 ± 13.69	48.71 ± 17.45	$< .001$
Currently smoking	599(62.3)	330(67.6)	267(69.9)	27(71.1)	0.006
FEV1 predicted, %	94.5 ± 12.6	92.0 ± 8.4	69.8 ± 7.2	43.5 ± 5.5	$< .001$
FEV1/FVC	0.75 ± 0.03	0.66 ± 0.03	0.61 ± 0.06	0.48 ± 0.08	$< .001$
TLV/pTLC	0.91 ± 0.12	1.01 ± 0.12	0.96 ± 0.12	0.99 ± 0.14	$< .001$
%LAA-910	12.42 ± 10.05	19.89 ± 12.07	18.99 ± 12.37	31.73 ± 12.75	$< .001$
%LAA-950	0.74 ± 0.89	1.66 ± 1.81	2.10 ± 2.95	6.69 ± 7.01	$< .001$

Table 6.1: Demographic, clinical and CT scan data by study group. Data presented as mean ± standard deviation or percentage in parenthesis. The p-values were estimed using χ^2-tests.

Gene-ration	Expected airways	No AFL (n = 961)	GOLD 1 (n = 488)	GOLD 2 (n = 382)	GOLD 3+4 (n = 38)	p-value
0	1	1.0 (1.0-1.0)	1.0 (1.0-1.0)	1.0 (1.0-1.0)	1.0 (1.0-1.0)	1
1	2	2.0 (2.0-2.0)	2.0 (2.0-2.0)	2.0 (2.0-2.0)	2.0 (2.0-2.0)	0.776
2	4	4.0 (4.0-4.0)	4.0 (4.0-4.0)	4.0 (4.0-4.0)	4.0 (4.0-4.0)	0.866
3	8	7.8 (7.0-8.0)	8.0 (7.0-8.0)	7.8 (7.0-8.0)	7.8 (7.0-8.0)	0.193
4	16	15.8 (15.0-16.0)	15.6 (14.8-16.0)	15.0 (14.0-16.0)	14.9 (13.2-15.8)	$< .001$
5	32	27.0 (24.8-29.0)	26.0 (23.0-28.0)	23.0 (19.0-27.0)	19.0 (15.0-25.9)	$< .001$
6	64	32.0 (24.4-37.8)	27.0 (20.0-34.0)	20.8 (13.0-27.0)	15.0 (7.2-19.0)	$< .001$
7	128	22.0 (15.0-28.8)	16.0 (11.0-23.0)	11.0 (6.0-16.8)	8.4 (5.0-11.9)	$< .001$

Table 6.2: Number of analysed airways by generation and study group. Data presented as median and interquartile range in parenthesis. The p-values were estimated using χ^2-tests.

Table 6.2 shows the number of analysed airways by AFL group and generation as mean values across all visits. A total of 195,199 airway branches were analysed, each measured at an average of 4.6 points in time. The number of airways that were detected in all time-points and could be analysed increases with generation until generation 5-6, where it starts to decrease. Fewer airways are found and analysed in participants with AFL as compared to subjects with no AFL; this effect is only small in generation 4, and increases from generation 5 and up.

The variation in inspiration level was considerable with a standard deviation of TLV/pTLC within each subject (mean ± standard deviation) of $4.8 \pm 2.9\%$. This is, however, far less variation than between in- and expiratory scans.

6.3.2 Airway distensibility and lumen diameter

In the mixed effects regression model all interactions between AFL, generation, and log(TLV/pTLC) were significant predictors of ALD ($p < 0.001$). There

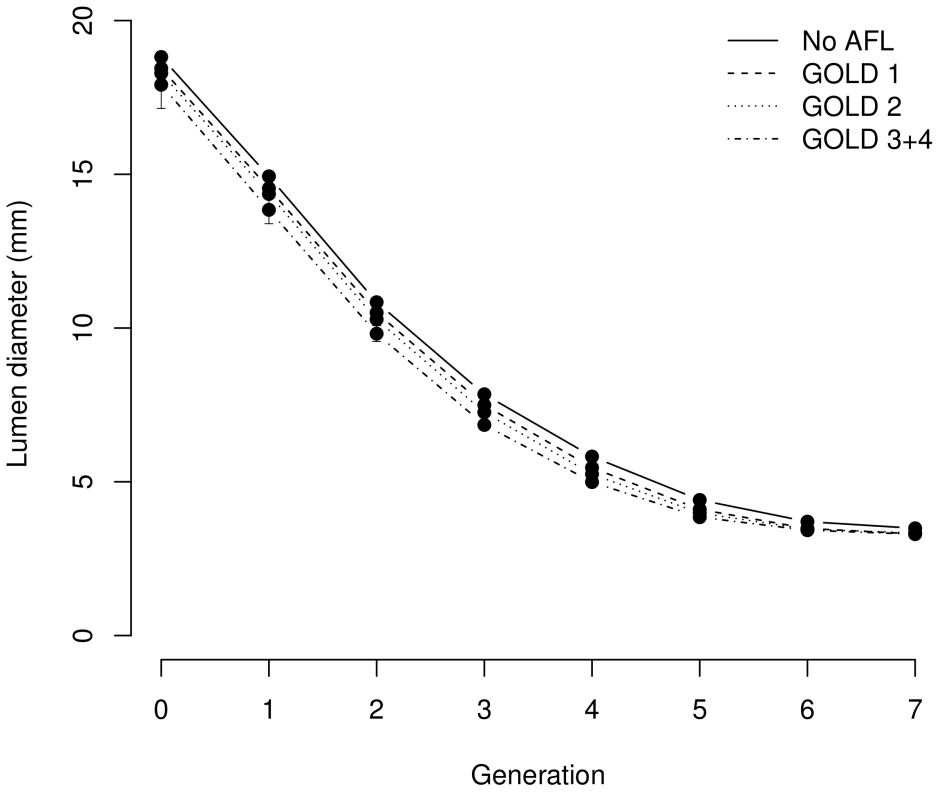


Figure 6.2: Lumen diameter (mm) in airway generations in No AFL and GOLD groups 1, 2 and 3+4, respectively.

was a significant ageing effect ($p < 0.001$), however interactions between age and the other explanatory variables were not found to improve predictions of log(ALD) ($p > 0.110$) and were thus removed from the model.

ALD in airways of the same generation were 1.4% ($p < 0.05$), 2.0% ($p < 0.01$), and 3.3% ($p < 0.001$) larger in the older quartiles (56-60, 60-64, and 64-73 years respectively) compared to the youngest quartile (51-56 years).

Figure 6.2 shows ALD of different generations for participants in the youngest age quartile at pTLC as predicted by the model. ALD decreased with generation and disease severity ($p < 0.001$).

Distensibility of airways in different generations and from participants with varying AFL is shown in Figure 6.3. Distensibility increased with airway generation and decreased with severity of AFL ($p < 0.001$). There was no measurable (i.e. statistically significant) distensibility of the airways in participants with severe AFL (GOLD 3+4, $p > 0.276$).

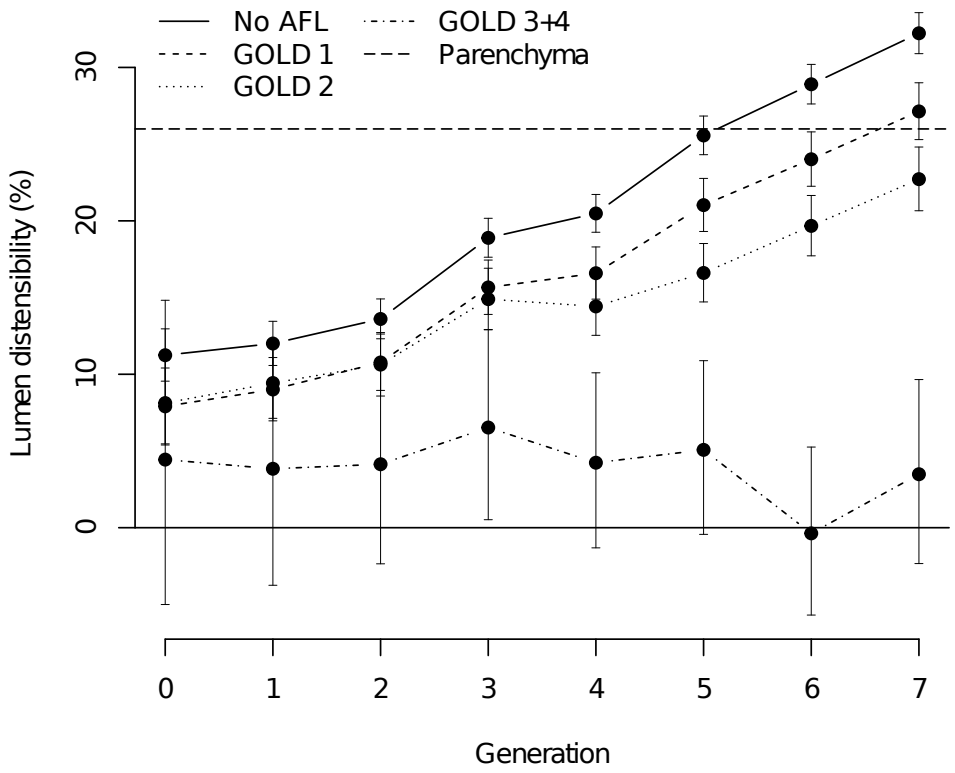


Figure 6.3: Lumen distensibility (%) in airway generations in the No AFL and GOLD groups 1, 2 and 3+4, respectively. Parenchyma line shows what distensibility has to be for airway volume to change relatively just as much as lung volume (assuming equal distension in all directions)

6.3.3 Airway wall thickness

In the mixed effects regression model all interactions between AFL, generation, and log(TLV/pTLC) were significant predictors of AWT ($p < 0.001$). Age and interactions between age and the other explanatory variables were not found to improve predictions of log(AWT) ($p > 0.061$) and were thus removed from the model.

Figure 6.4 shows AWT of different generations at pTLC as predicted by the model. AWT decreased with generation and increased with disease severity ($p < 0.001$).

6.4 Discussion

We observed several effects on large airways increasing with severity of COPD: A substantial reduction in the number of airways found and matched in all

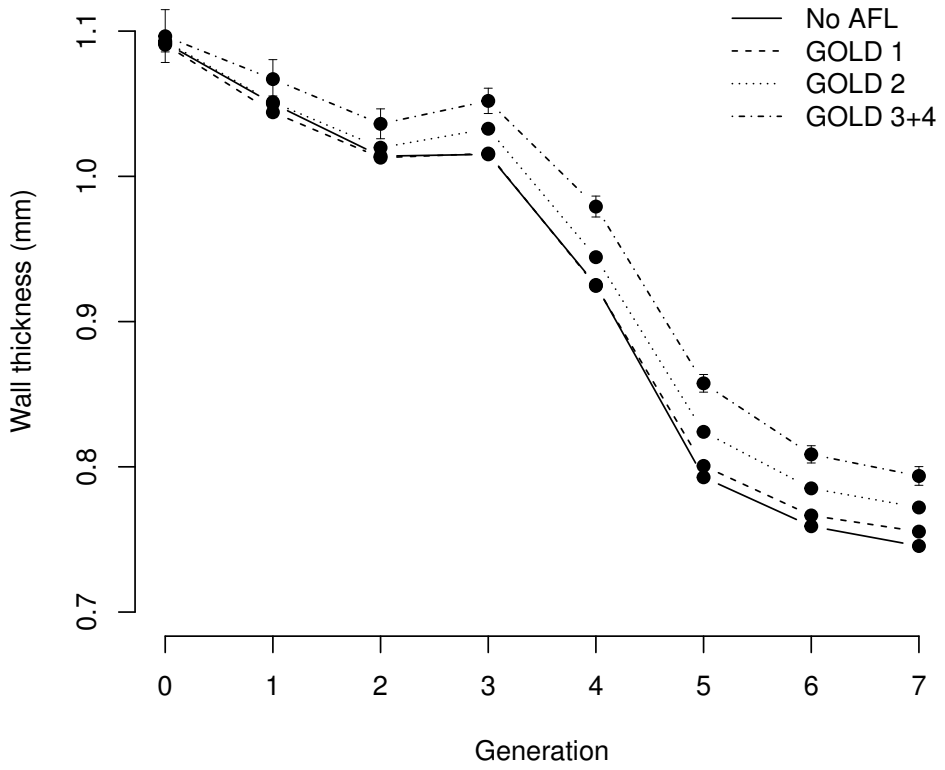


Figure 6.4: Wall thickness (mm) in airway generations in the No AFL and GOLD groups 1, 2 and 3+4, respectively.

consecutive scans in generations 5-7, a decrease in airway diameter, an increase in wall thickness, and a decrease in airway distensibility; in GOLD group 3+4 no distensibility was found at all.

6.4.1 Number of airways

Fewer airways are found in COPD, and this effect is increasing in higher generations and more severe disease. This was already observed more than 50 years ago in bronchographic studies (Hogg et al. 1968; Simon and Galbraith 1953; Freimanis and Molnar 1960). Possible explanations at that time were mucus plugging and emphysematous destruction of bronchi, and these reasons should also result in fewer detectable airways in CT images (Diaz et al. 2010a; Pu et al. 2012). Segmentation algorithms have problems penetrating luminal impactions and bridging emphysematous air spaces, however, airway narrowing should also move some airways, that would otherwise be visible, beyond the resolution of the CT images. These problems may impose a selection bias on airways that are segmented in COPD as compared to airways in subjects

with no AFL.

6.4.2 Airway lumen diameter (ALD), wall thickness (AWT), and distensibility

We observed decreased ALD and increased AWT in COPD, and this effect increased in higher generations and more severe disease. Interestingly, these results are similar to findings in previous studies of small airways; thus, histologic (Hogg et al. 1968; Hogg et al. 2004) and microCT (McDonough et al. 2011) studies have reported inflammatory changes in thickened airway walls and narrowing and disappearance of small conducting airways. The CT technique of the present study limits the segmentation of airways to large airways, i.e. bronchi lumen diameter is typically larger than 2 mm. Nevertheless, the trend toward more pronounced changes in higher airway generations lend support to a hypothesis proposed by others (Nakano et al. 2005) that changes in large airways may mirror changes in small airways.

We observed a markedly reduced distensibility of airways in patients with severe COPD, which is well-known from physiology (O'Donnell et al. 1987; Georgopoulos et al. 1993). Emphysema implies loss of airway attachments with loss of elastic recoil resulting in dynamic collapse of airways during expiration. Another important mechanism is airway inflammation, which through increasing thickness and fibrosis of the walls causes airways to become rigid and less distensible. This has been used to explain loss of distensibility when measured from exp- and inspiration scan-pairs (Diaz et al. 2012; Matsuoka et al. 2008). Our results show reduced distensibility even close to maximum inspiration, which shows dynamics are affected by COPD over a large range of inspiration levels. The clear separation of the COPD groups in terms of distensibility, observed in Figure 6.3, suggests that this reduced airway distensibility could be used as an important additional indicator for COPD.

If increased wall thickness in COPD is due to inflammation, one would expect that airway walls become thicker due to accumulation of inflammatory cells and oedema formation as compared to airways in healthy smokers with no AFL. But it is not possible to decide whether there is an inflammatory thickening of large airways with resulting lumen narrowing in our participants with severe AFL or if the observed thickening could simply be explained by reduced stretching of the airways during inspiration due to emphysema or increased airway rigidity.

In a previous study (Chapter 5) we observed that increasing age in healthy smokers is accompanied by larger airways (increased ALD). In COPD, distensibility is compromised, and increasing severity is accompanied by smaller lumen diameter, despite increasing age.

In addition to GOLD groups we could have used CT defined emphysema as a predictor in the regression models. But generally accepted categories of emphysema severity based on CT lung density have not yet been established.

The percentage of current smokers differs in the groups, but analysing current smokers and ex-smokers separately did not change conclusions.

6.4.3 Limitations

Our study has several limitations. The number of individuals with severe disease (GOLD group 3 and 4) in this study is quite limited. This is due to the fact, that the study population was recruited for lung cancer screening with an exclusion criterion, which ruled out those with low lung function, because in case of a positive screen with subsequent diagnosis of lung cancer, patients with low lung function (FEV1 $< 30\%$ predicted) could not be offered surgery.

There is a risk of selection bias because it is likely that the segmentation algorithm only finds the largest airways of generation 6 and 7, causing the average ALD to increase here, as indicated by the flattening at this point in the curve of Figure 6.2. Missing airway branches may, however, also have the opposite effect because airways distal to missing side branches are falsely classified as more proximal implying a reduced mean ALD. The overall effect of these opposite biases for various degrees of AFL is unknown.

The inspiratory level (TLV/pTLC) was not independent of the degree of AFL (table 1). The CT total lung volume (TLV) was smaller among participants without AFL as compared to GOLD group 1, which was probably due to hyperinflation in the COPD groups. However, there was no clear trend among the COPD groups, so these differences cannot explain the observed results, which do show a clear trend.

6.5 Conclusion

In this chapter we have studied the influence of COPD on the dynamic changes in the airways. We have adjusted for the confounding influence of the depth of inspiration when measuring the airway lumen diameter and airway wall thickness, and to our knowledge, this has not been done in previous studies.

In conclusion, airway distensibility decreases with the severity of COPD as defined by GOLD groups. Airway wall thickness increases with severity of COPD and airway lumen diminishes due to airway wall inflammation and/or lack of elastic recoil in surrounding emphysematous lung. Future analysis of emphysema, as detected by CT, and its effect on airway pathology may prove helpful in the process of understanding the interacting mechanisms in COPD.

Acknowledgements

The authors wish to thank Pechin Lo and Vladlena Gorbunova for developing initial versions of the software used for airway extraction and matching;

Piet L.B. Bruijnzeel, and Laura Hohwü Thomsen for their valuable contributions and the Netherlands Organisation for Scientific Research (NWO) and AstraZeneca Sweden for financial support.

Appendix

Statistical methods supplement

The statistical models take the form:

$$Y_{s,b,v} = \sum_{c\in\{1,2,3,4\}} \sum_{g\in\{0,1,\ldots,7\}} \Bigg(\beta_{0,c,g} \times C_s \times G_{s,b,g} + \beta_{1,c,g} \times C_s \times G_{s,b,g} \times \log\left(\frac{\mathrm{TLV}_{s,v}}{\mathrm{pTLC}_s}\right)\Bigg) + \sum_{a\in\{1,2,3,4\}} (\beta_{2,a} \times A_{a,s}) + \alpha_s + \alpha_{s,b} + (\gamma_s + \gamma_{s,b}) \times \log\left(\frac{\mathrm{TLV}_{s,v}}{\mathrm{pTLC}_s}\right) + \epsilon_{s,b,v},$$

where the index s refers to the subject, b to the branch within the subject, and v to the visit. $c \in \{1, 2, 3, 4\}$ refers to GOLD group, $a \in \{1, 2, 3, 4\}$ to the age group, $g \in \{0, 1, \ldots, 7\}$ to the generation. The outcome variable Y is either log(ALD) or log(AWT). $\beta_{0,c,g}$, $\beta_{1,c,g}$, and $\beta_{2,a}$ are the coefficients of interests, describing the outcome variable's relationship with the interaction of GOLD group and generation; interaction of GOLD group, generation and inspiration level; and age of the subject. The random intercepts α_s and $\alpha_{s,b}$ are assumed to be independent, normally distributed with zero mean across subjects and within subjects, respectively. The residuals $\epsilon_{s,b,v}$ follow independent zero-mean normal distributions. The categorical variables $G_{s,b,g}$, $C_s \in \{0, 1\}$ and $A_{a,s} \in \{0, 1\}$ describe whether a branch in a subject belongs to a specific generation, whether a given subject is in a given GOLD group, and whether a given subject belongs to specific age group. The models are fitted using restricted maximum likelihood via the lme4 package (Bates et al. 2012) for the R programming language (R Core Team 2012).

Statistical results supplement

The standard deviation of the mean ALD in a generation was 14.3% between subjects belonging to the same age and GOLD group and with the same inspiration level. For AWT the corresponding number was 3.6%. The intra-subject standard deviation of ALD, standardizedfor inspiration level, was 28.2% in a given generation. For AWT the corresponding number was 8.4%.

The standard deviation of the mean lumen distensibility of airways in the same generation was respectively 34.4% between subjects of the same age and GOLD group. In terms of wall thinning the corresponding was 7.3%. The

intra-subject standard deviation of lumen distensibility in each generation was 9.7%. For wall thinning the number was 8.3%.

The standard deviation of the residuals was 6.6% for ALD. For AWT the number was 10.4%.

Chapter 7

Longitudinal segmentation using groupwise registration and 4D optimal surfaces

The work presented in this chapter is based on J. Petersen, M. Modat, M. J. Cardoso, A. Dirksen, S. Ourselin, and M. de Bruijne (2013c). "Quantitative Airway Analysis in Longitudinal Studies using Groupwise Registration and 4D Optimal Surfaces". In: *Med Image Comput Assist Interv - MICCAI 2013*. Ed. by K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab. Lecture Notes in Computer Science. Springer, pp. 287–294.

Abstract

Quantifying local changes to the airway wall surfaces from computed tomography images is important in the study of diseases such as chronic obstructive pulmonary disease. Current approaches segment the airways in the individual time point images and subsequently aggregate per airway generation or perform branch matching to assess regional changes. In contrast, we propose an integrated approach analysing the time points simultaneously using a subject-specific groupwise space and 4D optimal surface segmentation. The method combines information from all time points and measurements are matched locally at any position on the resulting surfaces.

Visual inspection of the scans of 10 subjects showed increased tree length compared to the state of the art with little change in the amount of false positives. A large scale analysis of the airways of 374 subjects including a total of 1870 images showed significant correlation with lung function and high reproducibility of the measurements.

7.1 Introduction

Assessing the dimensions and attenuation values of airway walls from Computed Tomography (CT) images is important in the investigation of airway remodelling diseases such as Chronic Obstructive Pulmonary Disease (COPD) Hackx et al. 2012. Manual measurements are very time consuming and subject to intra- and inter-observer variability. Automatic methods are needed to estimate the dimensions of large parts of the airway tree. Obtaining reproducible measurements is difficult because of a strong dependence on position in what is a complicated biologically and dynamically varied tree-like structure. Previous approaches have solved the problem in two steps: first a step, to segment the airways and conduct the measurements, and second a step to identify anatomical branches or match individual airway segments in multiple scans of the same subject Tschirren et al. 2005b; Petersen et al. 2011a; Feragen et al. 2012. The task of anatomically identifying airway branches poses significant problems even to medical experts Feragen et al. 2012. Most automatic methods therefore do not go beyond the segmental level, resulting in 32 labelled branches Feragen et al. 2012; Tschirren et al. 2005b. State of the art segmentation methods can extract many more branches reliably Lo et al. 2012, and intra-subject branch matching can thus increase the information available in longitudinal studies. It has been done using image registration Petersen et al. 2011a, or association graphs Tschirren et al. 2005b.

A limitation of such two-step approaches is that longitudinal information is not used to the fullest. For instance branches need to be segmented in every scan in order to be matched even though a branch that is detected in one scan is most likely present in all. The proposed method improves on this by segmenting multiple scans of the same subject simultaneously. It is thus able to combine information from all time points and enables matched measurements, not just at the branch level, but locally at any point on the resulting surfaces.

7.2 Methods

An initial airway lumen probability map (section 7.2.2) was constructed by transferring initial segmentations of each time point to a per-subject common space constructed through deformable image registration (section 7.2.1). A four dimensional optimal surface graph was built from the initial probability map and used to find the inner and outer airway wall surfaces as the global optimum of a cost function combining image terms with surface smoothness, surface separation, and longitudinal penalties and constraints (section 7.2.3).

7.2.1 Groupwise registration of images

Prior to registration, intensity inhomogeneity due to for instance gravity gradients and ventilation differences was removed within the lungs using NiftySeg (`http://sourceforge.net/projects/niftyseg`) Cardoso et al. 2011. The approach is using a two-class expectation-maximization based probabilistic framework, which incorporates both a Markov Random field spatial smoothness term and an intra-class intensity inhomogeneity correction step.

All registrations have been performed with a stationary velocity field parametrisation and using normalised mutual information as a measure of similarity with NiftyReg (`http://sourceforge.net/projects/niftyreg`) Modat et al. 2010. For each subject, all time points were aligned to the Frechet mean on the space of diffeomorphisms, thus providing a common space for analysis and a one-to-one mapping between time points.

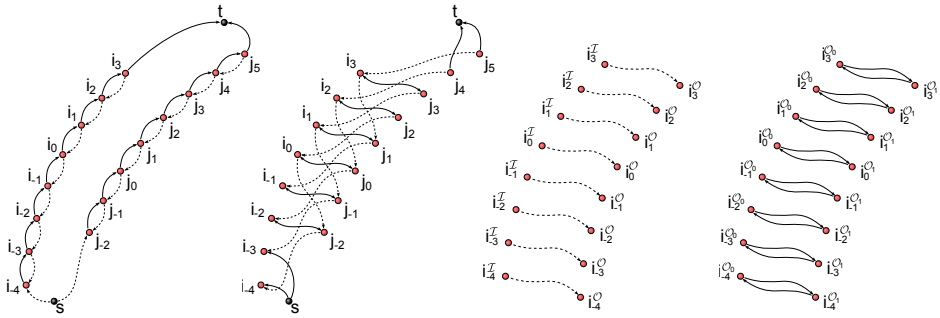
7.2.2 Initial segmentation

An initial airway lumen segmentation was obtained in each individual image using the Locally Optimal Path (LOP) approach of Lo et al. 2009. Segmentations from every scan of the same subject were then warped to the subject-specific groupwise space and averaged, giving a lumen probability map. By thresholding this using some value T, it is possible to move freely between the intersection and union of the segmentations, weighting the amount of included branches against the amount of false positives. Disconnected components were connected along centrelines extracted from the union segmentation using the method described in Lo et al. 2012. The voxel based initial segmentation was then converted to triangle mesh with vertices $\mathcal{V}$ and edges $\mathcal{E}$, using the marching cubes algorithm.

7.2.3 Graph construction

This section describes how a graph $G = (V, E)$ can be constructed, such that the minimum cut of G defines the set of surfaces $M = \mathcal{I} \cup \mathcal{O}$ where $\mathcal{I} = \{\mathcal{I}_0, \mathcal{I}_1, \ldots, \mathcal{I}_N\}$ and $\mathcal{O} = \{\mathcal{O}_0, \mathcal{O}_1, \ldots, \mathcal{O}_N\}$ are the inner and outer surfaces of the N scans of the subject. The graph is similar to that of Petersen et al. 2011b, the differences being the addition of the longitudinal connections and hard constraints.

The vertices of the graph are defined by a set of columns V_i^m one for each vertex $i \in \mathcal{V}$ of the initial mesh and sought surface $m \in M$, and a source s and a sink vertex t. As in Petersen et al. 2011b we will let the columns be defined from sampled flow lines traced inward and outward from $i \in \mathcal{V}$. Because they are non-intersecting, the found solutions are guaranteed to not self-intersect. Tracing is done within a scalar field arising from the convolution of the binarised initial segmentation with a Gaussian kernel of scale σ. Sampling is done at regular arc length intervals relative to i until



(a) Data and topology (b) Smoothness (c) Separation (d) Longitudinal

Figure 7.1: Example columns with $I_i = 4$, $I_j = 2$, $O_i = 3$ and $O_j = 5$ inner and outer vertices illustrating the graph as implemented. The dotted edges have infinite capacity and implement hard topology, smoothness and separation constraints. The solid edge capacities are given by data term, smoothness, separation and longitudinal penalties.

tracing is stopped due to flattening of the gradient, giving I_i and O_i inner and outer column points. So, $V_i^m = \{i_k^m \mid k \in K_i\}$, where $K_i = \{-I_i, 1 - I_i, \ldots, 0, \ldots, O_i\}$ and $V = \bigcup_{i \in \mathcal{V}, m \in M} V_i^m \cup \{s, t\}$. In this way, the column defines the set of all possible solutions for i in the surface m.

Let $(v \xrightarrow{c} u)$ denote a directed edge from vertex v to vertex u with capacity c and $w_i^m(k) \geq 0$ be a cost function, giving the cost of vertex k in a column V_i^m being part of the surface m. This data term can be implemented by the edges:

$$E_d = \Big\{ \big\{ (i_k^m \overset{w_i^m(k)}{\rightarrow} i_{k+1}^m) \mid k, k+1 \in K_i \big\} \cup \big\{ (i_{O_i}^m \overset{w_i^m(O_i)}{\rightarrow} t), (s \overset{\infty}{\rightarrow} i_{I_i}^m) \big\} \mid i \in \mathcal{V}, m \in M \Big\} . \tag{7.1}$$

To prevent degenerate cases where a column is cut multiple times and to preserve topology, the following infinite cost edges are added:

$$E_\infty = \Big\{ (i_k^m \overset{\infty}{\rightarrow} i_{k-1}^m) \mid \ i \in \mathcal{V}, m \in M, k-1, k \in K_i \Big\} . \tag{7.2}$$

An example of these edges is given in figure 7.1a.

Let $f_{i,j,m,n}(|k-l|)$ be a convex non-decreasing function giving the pairwise cost of both $i_k^m \in V_i$ and $j_l^n \in V_j$ being part of the surfaces $m, n \in M$ respectively. This can be used to implement surface smoothness and longitudinal penalties, see (equation 7.7). Additionally let the set $I(i_k^m, j, n) = \{\zeta, \zeta + 1, \ldots, \eta\} \subseteq \{-I_j, 1 - I_j, \ldots, O_j\}$ put pairwise constraints on the solution, such that if i_k^m and j_l^n are both part of it, then $l \in I(i_k^m, j, n)$. Such

penalties and constraints can be implemented by:

$$\begin{aligned} E_i = \Big\{ & \big\{ (i_k^m \overset{\triangle(i_k^m, j_l^n)}{\rightarrow} j_l^n) \mid k \in K_i, l \in K_j \big\} \cup \\ & \big\{ (s \overset{\triangle(i_k^m, j_l^n)}{\rightarrow} j_l^n) \mid l \in K_j, k \in K_j, k < -I_i \big\} \cup \\ & \big\{ (i_k^m \overset{\triangle(i_k^m, j_l^n)}{\rightarrow} t) \mid k \in K_i, l \in K_i, l > O_j \big\} \\ & \mid i, j \in \mathcal{V}, m, n \in M \Big\} , \end{aligned} \tag{7.3}$$

and $\triangle$ gives the capacity of the edges as follows:

$$\triangle(i_k^m, j_l^n) = \begin{cases} \infty & \text{if } l = \min I(i_k^m, j, n) \\ 0 & \text{if } l \notin I(i_k^m, j, n) \\ \hat{\triangle}(k-l) & \text{otherwise} \end{cases} \tag{7.4}$$

where

$$\hat{\triangle}(x) = \begin{cases} 0 & \text{if } x < 0 \\ f_{i,j,m,n}(1) - f_{i,j,m,n}(0) & \text{if } x = 0 \\ f_{i,j,m,n}(x+1) - 2f_{i,j,m,n}(x) + f_{i,j,m,n}(x-1) & \text{if } x > 0. \end{cases} \tag{7.5}$$

Similar to the approach of Liu et al. 2012, hard constraints was used to force the outer surfaces to be outside their corresponding inner surfaces and solutions to not vary more than γ and δ indices in neighbouring columns in the inner and outer surfaces as follows:

$$I(i_{k,m}, j, n) = \begin{cases} \{k, k+1, \ldots, O_j\} & \text{if } m \in \mathcal{I}_s \text{ and } n \in \mathcal{O}_s, \\ & s \in \{0, 1, \ldots, N\} \\ \{k-\gamma, k-\gamma+1, \ldots, k+\gamma\} & \text{if } m = n, \text{ and } n, m \in \mathcal{I} \\ \{k-\delta, k-\delta+1, \ldots, k+\delta\} & \text{if } m = n, \text{ and } n, m \in \mathcal{O} \\ K_j & \text{otherwise.} \end{cases} \tag{7.6}$$

The following pairwise cost were implemented:

$$f_{i,j,m,n}(x) = \begin{cases} p_m x & \text{if } m = n \text{ and } (i,j) \in \mathcal{E} \\ qx & \text{if } i = j \text{ and } m \in \mathcal{I}_s, n \in \mathcal{I}_{s+1} \text{ or } m \in \mathcal{O}_s, n \in \mathcal{O}_{s+1} \\ 0 & \text{otherwise.} \end{cases} \tag{7.7}$$

p_m is the smoothness penalty, defining the cost of each index the solution varies between neighbouring columns in the same surface m. q is the longitudinal penalty, defining the cost inherent in the solution for each index corresponding surfaces are separated in each column within the groupwise space. An illustration of these edges is given in figure 7.1b and 7.1c. The total edge set E is given by: $E = E_d \cup E_\infty \cup E_i$.

The cost functions $w_k^m(k)$ were set to the first order derivative of the image intensity in the outward and inward direction of the flow line for the inner and outer surfaces respectively.

We used the algorithm described in Boykov and Kolmogorov 2004 to find the minimum cut.

7.3 Experiments and results

7.3.1 Data

CT images and lung function measurements from the Danish lung cancer screening trial Pedersen et al. 2009 were used. Images were obtained using a Multi Detector CT scanner (Philips Mx 8000) with a low dose (120 kV and 40 mAs), reconstructed using a hard kernel (D) with a resolution of approximately 0.78 mm $\times$ 0.78 mm $\times$ 1 mm. The subjects (at inclusion) were men and women, former and current smokers with at least 20 pack years smoked, between 50 and 70 years of age and thus at high risk of having COPD.

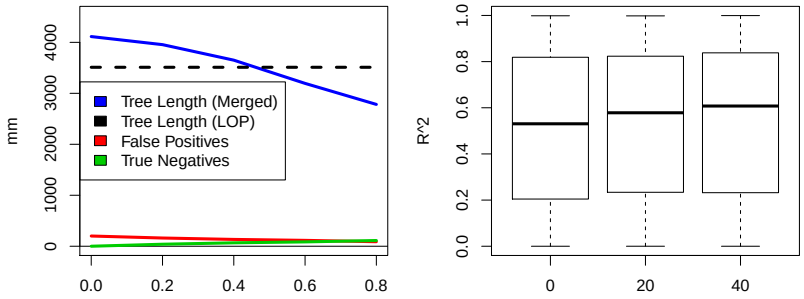
We used images of 10 subjects for estimation of the parameters and evaluation of the process of merging the initial segmentations. An independent set of 374 subjects was used to evaluate the ability of the method to detect longitudinal changes in airway dimensions. From each subject 5 yearly scans were included out of which 1739 had matching lung function measurements.

7.3.2 Parameters and merging of initial segmentations

The centrelines of the 10 subjects were moved from the groupwise space to the space of the last time point scan, in which they were manually checked using in-house developed software. The software allows movement along the centrelines while displaying a cross-sectional view of the airway.

Figure 7.2a and 7.3a show results of varying T - significantly more branches can be found by fusing information from multiple scans ($p < 0.05$ for $T < 0.4$), while the percentage of false positives does not seem to increase much. We chose a value of $T > 0.2$ corresponding to branches present in at least two scans.

Parameters were estimated by aiming to penalize and constrain solutions as little as possible while still preventing noisy segmentations and the inclusion of abutting vessels. The mesh edges were roughly 0.4 mm apart and the flow lines were sampled at 0.4 mm spacings. p_m was set to 15 for all $m \in M$, γ and, δ to 2 and σ to 0.7 based on visual inspection of results. Figure 7.3b shows a segmentation result illustrated at two time points.



(a) Threshold (b) Longitudinal penalty

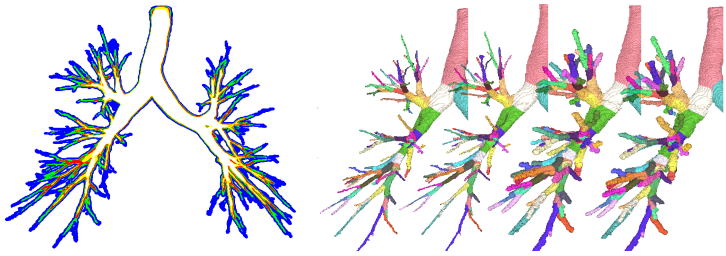
Figure 7.2: Left: Average tree length for different values of the threshold T (blue), average tree length of the LOP segmentation as a comparison (black dashed), false positives (red) and true negatives (green) with respect to the union of the segmented branches. Right: the R^2 fit of the relative change in lung volume predicted from relative change in lumen diameter in each subject with increasing amount of longitudinal penalty.

7.3.3 Longitudinal segmentation

Experiments were conducted on the images of the 374 subjects with values $q \in \{0, 20, 40\}$ of the longitudinal penalty to assess its impact on the method's ability to detect changes in airway dimensions over time. Note that a value of 0 is similar to performing multiple independent (3D) searches, but with the added bonus that the solution meshes will have corresponding vertices. Measures of Wall Area percentage (WA%) and Lumen Diameter (LD) were computed from the average distance of the mesh vertices to the branch centreline, which was extracted from the average segmentation in the groupwise space and transformed to each individual time point. This enables accurate assessment of subtle localized changes in airway morphology, but to evaluate the performance of our segmentation to automatically derive known airway imaging biomarkers, we averaged the measures extracted in the airways of generation 3 to 6.

WA% was found to correlate significantly with lung function at each time point (Average Spearman's ρ: -0.29 ± 0.01, $p < 0.0001$). Different values of q did not change the results significantly. Reproducibility, quantified by correlating results at time point one with those of time point two with $q = 0$, was high: (Spearman's ρ: 0.95, 0.94, 0.89 and 0.85 in generation 3, 4, 5 and 6 respectively $p < 0.0001$). As expected, increasing q made the reproducibility go towards 1.

Annual changes of WA% were found to be 0.31 ± 4.9 % significantly different from 0 (Mann-Whitney U test $p < 0.001$) with $q = 0$. Significance disappeared with $q \in \{20, 40\}$ perhaps evidence that these values are over-penalizing the solutions. No correlation was found between annual changes in



(a) (b)

Figure 7.3: Left: initial segmentation example, at $T > 0$ (blue), $T > 1/5$ (green), $T > 2/5$ (red), $T > 3/5$ (yellow), and $T > 4/5$ (white). Right: inner followed by outer surface segmentations at two corresponding time points. Colours show matching branches.

WA% and annual changes in lung function. This is not surprising giving the relative slow developing nature of COPD and the known poor reproducibility of lung function measurements. It is similar to what was previously reported Petersen et al. 2011a.

LD is dependent on the inspiration level and investigating the method's ability to detect this dependency can therefore be used as a surrogate for a much slower pathological change. The relative change in lung volume was thus predicted from the relative change in LD in each subject. The R^2 of all the models showed significantly better fit when using higher values of q ($p < 0.0001$ using Wilcoxon signed rank test), see figure 7.2b, indicating that the chosen longitudinal penalties can improve the ability to detect changes in the inner surface.

7.4 Discussion and conclusion

We have presented a method for the analysis of airways designed to fully exploit longitudinal imaging data. The method, in contrast to state of the art, can use information from all time points and because it outputs surface meshes with one-to-one correspondences between vertices, enables measurements to be compared locally without the need for a separate branch matching step. A visual evaluation of the scans of 10 subjects showed the method found significantly more complete airway trees with minimal changes to the false positive percentage. Results on 1870 scans show significant correlation with lung function and highly reproducible results. Future work will have to better investigate choices for the longitudinal penalty. For instance it is possible, as indicated by the experiments, that different values are needed for the inner and outer surfaces, due to the different contrast values between lumen and wall and between wall and lung parenchyma. It should also be noted that

averaging measurements over large parts of the airway tree, as done in this work, is ignoring information provided by the matched measurements, however to limit the scope of the paper we left it to future work to develop statistical models incorporating this information. Such models should, we expect, have more power to detect changes.

Acknowledgements.

This work was partly funded by EPSRC (EP/H046410/1), the National Institute for Health Research University College London Hospitals Biomedical Research Centre (Award 168), the Netherlands Organisation for Scientific Research (NWO), and AstraZeneca Sweden.

Chapter 8

Summary and general discussion

8.1 Summary

Chapter 2 presented a new curvilinear column construction method for optimal surface methods (Wu and Chen 2002). The method enables solution surfaces that are guaranteed to not self-intersect and is well suited for high-curvature regions. The proposed method was applied to segment airway walls from low-dose CT images, showing sub-voxel accuracy on phantom scans, and significantly higher overlap with manual annotations and visually improved results compared to a similar approach using straight columns (Liu et al. 2012). Airway abnormality measurements obtained with the method are reproducible and correlate significantly with lung function.

Chapter 3 presented a method, which matches individual branches in multiple scans of the same subject. The method allows changes in dimensions of individual airway branches to be tracked from scan to scan, which could increase the statistical power of longitudinal studies. The presented results show that variability of measurements due to missing and spurious branches can be reduced, by limiting measurements to consistently found branches. The results of later studies presented in Chapter 5 and 6 also show that the method is useful to investigate airway dynamics.

Chapter 4 presented a supervised algorithm for labelling of airway trees, which assigns anatomical labels of the airway branches down to the segmental level using geodesic distances in a geometric tree-space. The algorithm allows measurements in the same anatomical branches to be compared from subject to subject, which removes some of the variability due to not measuring the same position within the airway trees. The algorithm was evaluated on 80 segmented airway trees from 40 subjects at two time-points, each labelled by three medical experts, showing the algorithm to be statistically similar to the experts in accuracy and better in terms of reproducibility. The performance of

the algorithm was determined, using a very large data set of more than 8000 segmented airway trees, to not depend on disease severity, assuming it is given equally complete airway trees. Results of the thorough evaluation presented provide insights into not only performance of the algorithm but also that of medical experts (Petersen et al. 2013b).

The methods of Chapter 2, 3, and 4 have allowed us to analyse the unique data set of the Danish Lung Cancer Screening Trial (DLCST) (Pedersen et al. 2009). Access to the data gave us a unique opportunity to study properties of airways in a group of subjects where almost half were diagnosed with COPD.

Chapter 5 presented a study into the effect of inspiration level at time of scan on measured airway dimensions in subjects with normal lung function. Even when subjects are asked to inspire maximally variations in inspiration level still exist (an intra-subject standard deviation of 5.4% was observed). The distensibility of airways means that measured airway dimensions are affected by these variations. The presented study shows, for the first time, the size of this effect by generation and airway segment. It shows that subjects who inspire deeper prior to scanning tend to have larger lumen diameter and thinner walls, an effect which is more pronounced in more peripheral airways. It also shows that airways in generation 5 and above, roughly corresponding to the segmental level and beyond, are as distensible as the lung parenchyma. This suggests that adjustment for variation due to differences in inspiration level is needed and that adjustment may be relatively straight-forward in peripheral airways, as the volume of these airways change the same amount relatively as the lungs. Adjustment of for instance the lumen diameter may thus be done by dividing by the cubic root of the inspiration level, assuming inspiration level is quantified as was done in Chapter 5 by the total lung volume divided by the predicted total lung capacity.

Chapter 6 presented a study into airway dimensions and distensibility in subjects with COPD. Statistical models were used to investigate the effects of inspiration level, disease severity, age and measurement position on lumen diameter and wall thickness. The results show that distensibility decreases, wall thickness increases and airway lumen diminishes with severity of COPD and these changes are more pronounced toward smaller airways. These changes indicate that measured airway dimensions are both statically and dynamically influenced by severity of COPD. This is possibly due to airway wall inflammation and/or lack of elastic recoil in surrounding emphysematous lung.

Chapter 7 presented a method, which should be more suited to measure differences in the airways from scan to scan, allowing effects of differences in inspiration level and longitudinal change to be more precisely determined. Rather than segmenting each time point individually and then matching the branches to assess changes, the presented approach can segment all scans of the same subject simultaneously. This allows the method to combine information from multiple images, which results show can increase the length of the segmented airway tree with little change in the amount of false positives.

The method is well suited to measure longitudinal or dynamic change because it outputs corresponding surface meshes, meaning change due to pathology or differences in inspiration level can be measured locally in any point on the surfaces, without the need for a separate branch matching step.

8.2 General discussion

Diagnosis of COPD is based on spirometry, however, spirometry may be insensitive to early changes (Hackx et al. 2012), has only a weak correlation with symptoms and health status (GOLD 2013; Jones 2009), and provides little information on the underlying causes for limitation of airflow seen in COPD. It is known that two factors play a role, emphysema and small airway disease. The contributions of these factors is unclear and further research is needed to establish whether they can be viewed as independent disease phenotypes (Decramer et al. 2012). CT and modern image analysis tools, such as those presented in this thesis, play an important role in investigating possible phenotypes as they allow quantification of both emphysema and airway abnormalities. Phenotypes could be important in predicting clinical characteristics and responsiveness to treatment with bronchodilators and inhaled corticosteroids (Kitaguchi et al. 2006; Fujimoto et al. 2006). There is also evidence that the presence and distribution of emphysema (Martinez et al. 2006; Johannessen et al. 2013) and the interaction of airway wall thickness and emphysema (Johannessen et al. 2013) affects mortality. Airway analysis from CT images could thus potentially be used in patient care, early clinical trials to determine drug efficacy or input to prognostic models to predict patient outcomes. Although many challenges remain to make most of this a reality, the work presented in this thesis shows that fully automated airway analysis of large data sets is possible and that it can be used to detect abnormalities significantly even in mild COPD cases from low-dose images.

Chapter 2 presented a graph based method for airway wall segmentation and Chapter 7 presented an extension of it to joint segmentation of multiple images. An advantage of these methods is that they are able to extract all surfaces simultaneously in an optimal fashion, which means that they can use positions of each surface to help place all other. A disadvantage of these methods is that they are discrete, which means that every possible position the sought surfaces can take needs to be represented as a node in the graph. The methods, therefore, require relatively large amounts of memory and a compromise between memory usage and resolution often has to be made. We chose to place the nodes of the graph at approximately 0.5 mm intervals, which is not far from the average thickness of the airway walls (Figure 6.4) and so this graph resolution may be too rough to detect some of the more subtle changes to the wall caused by COPD and differences in inspiration level. This relative low resolution graph was chosen in order to allow the extensive parameter

tuning of Section 2.3.2 to be conducted in a reasonable time-frame. The relative low resolution graph thus made it possible to compare methods based on well fitted parameters, however, better segmentation results can probably be obtained with a higher resolution graph with an acceptable increase in memory usage.

The results presented in Chapter 5 on the relationship between wall thickness and differences in inspiration level show the airway wall to be thinning as it is stretched during inspiration. This is not an unexpected result, but as far as we are aware it is the first time it has been measured. The results also seem to indicate that wall volume is not constant but actually increasing during inspiration. As we are not aware of any physiological process which could increase the volume of the wall during inspiration, we chose to not recommend an adjustment for the effect. It is possible the low resolution of the graph is to blame for this unexpected finding, as some of the changes to the wall could be too small to be seen by the method. The low-dose CT images and partial volume effects could, however, also play a role. It would be interesting to repeat the experiments with a higher resolution graph and if possible higher resolution images. Such new results would perhaps agree more with a hypothesis of wall volume preservation under stretching and such experiments could therefore allow us to derive a simple approach to adjust for inspiration level differences.

Airway dimensions change with location and different strategies to obtain comparable measurements have been explored in the literature. A common strategy is to identify and measure the same anatomical bronchi in each scan (Nakano et al. 2000; Lederlin et al. 2012; Diaz et al. 2012; Hasegawa et al. 2006; Brown et al. 2001). The process of labelling these branches from segmented airway trees can now be automated (Ginneken et al. 2008b; Lo et al. 2011; Tschirren et al. 2005b). The labelling method presented in Chapter 4 is one such automated approach. Unfortunately, as revealed by the results of that chapter, labelling is difficult, even for medical experts. For instance, R6 and L6 are the only segmental bronchi to be identified by the experts with more than 90 % accuracy and 9 of the segmental bronchi are identified with accuracies lower than 70%. Despite the fact that the presented algorithm is as accurate as the medical experts, one should question when labelling is a good strategy and where in the airway tree it can be used to reduce variability of measurements. Figure 5.2 and 5.3 shows that there is a clear relationship between properties such as lumen diameter, wall thickness, distensibility and generation. Segmental bronchi, however, seem to be spread out over the range of values represented by the airways at generation 3 to 7. This means that the segmental level itself is not necessarily indicative of certain properties and measurements conducted in different segmental bronchi should therefore not be compared without reservations. Airways can be assumed to branch by dichotomy and this means that in a normal airway tree, branches of each generation should be progressively and predictably smaller (Weibel 2009). Gen-

erations are also much easier to detect than anatomical labels and all of the branches of the segmented airway tree can be assigned a generation and consequently analysed. Moreover, it is possible to estimate the completeness of the segmented airway tree in each generation, because the number of branches should double with each bifurcation (Figure 5.1). Grouping measurements by generations as opposed to by anatomical label therefore seem to be a more useful and easy to implement strategy. Because many branches contribute to a measurement, some of the noise present in measuring each individual branch may also be evened out, resulting in a better assessment of the average state of the airways in each individual subject. The overall smaller standard errors of the estimates of generation 1 to 7 compared to each segmental bronchus seen in Figure 5.2, 5.3, and 5.4 is evidence to support this. Measurements should, however, not be compared without reservations in generations that are not nearly completely found, as the subset of airways found in such generations are likely the most easily visible and therefore not representative. Although anatomical labels may be less useful than generations for comparing measurements between subjects, they can be used to provide additional information, for instance by using a combination of generations and named bronchi to study distribution of airway abnormalities by position within the lung. This could be done by counting generations from each of the lobar or the subset of the segmental bronchi that can be identified with greatest accuracy.

Chapter 5 and 6 show disease, distensibility and variations in inspiration level affect measurements of airway dimensions in maximum inspiration scans. Subjects who inspire more during scanning will appear to have thinner airway walls and wider airway lumen, that is, they will appear to have less small airway disease. The opposite is true for density measurements of emphysema, where subjects who inspire more will appear to have less dense lungs and therefore more emphysema. It is thus questionable whether it is possible, as previous studies have attempted (Dijkstra et al. 2013; Nakano et al. 2000), to evaluate the independent contribution of airway abnormalities and emphysema to airflow limitation without considering the effects of variations in inspiration level. The methods presented in this thesis could be used to conduct such an investigation.

8.3 Future prospects

The airway wall segmentation method presented in Chapter 2 used a weighting of the first and second order derivatives of the image intensity to find the position of the airway wall surfaces. Simple intensity based surface terms like this are common in airway wall segmentation (Li et al. 2006; Liu et al. 2012; Petersen et al. 2010) but for airway extraction (and also other segmentation tasks), inclusion of higher order texture information, for instance by using supervised learning to build voxel appearance models, have resulted in improved

results (Lo et al. 2010). Such models can be used to generate probabilities of voxels being lumen, wall or background and these probabilities could then be integrated into the segmentation method of Chapter 2 using the regional terms of (Haeker et al. 2007). Regional terms, loosely said, not only consider the position of the surface, but also what is inside them. One reason this is an advantage is that outside structures, such as nearby vessels, are often more clearly defined than the walls of especially smaller airways, meaning that with the implemented surface terms they can be erroneously included. Regional terms, however, consider the complete region between the surfaces and these outside structures will thus only be included if all of the in-between area looks like wall, which should be less likely. Tuning such a method with phantoms could be more challenging, however, as it will be able to recognise more subtle signs of the tissue it is trained to classify. Simple plastic tubes without something resembling parenchyma and nearby vessels may thus not be enough.

Experiments indicate that the presented wall segmentation method (Chapter 2) may be able to grow complete airway trees if it is iterated multiple times, for instance by starting from an initial spherical seed surface placed within the trachea. The advantage of such an approach compared to region growing is that it grows by looking for both the inner and outer airway wall surface simultaneously and takes smoothness into account; continued growth is therefore not certain apon leakage into the parenchyma. Although no guarantees are given on the amount of iterations it takes to complete, in practice it seems to move towards the final solution in a steady fixed pace. This pace is generally lower with higher smoothness constraints/penalties and bounded above by the length of the flow lines. This nice behaviour can be explained by the fact that it only grows if the new solution is determined to be more optimal (Equation 2.6) than the previous iteration. So it is not possible for the method to return to a solution of a previous iteration and thereby iterate endlessly. There are however problems, which remain to be solved. Re-meshing is needed in regions that grow in order to keep the mesh vertices evenly distributed. Re-meshing, however, changes the optimality of a given solution, meaning it makes it possible for the method to jump back to an already visited solution. Another problem occurs because smoothness is measured relative to the previous iteration, that is, the previous iteration is always considered perfectly smooth. This makes the method biased towards not growing and probably stops growth early in some cases. If these problems can be solved or the consequences of them lessened, this new iterative method could remove the need for an initial segmentation and allow airways to be more completely segmented with less leakage.

The methods of Chapter 2 and 7 were developed to segment airway walls from CT. Since then the method of Chapter 2 has also been applied to segment carotid artery walls from magnetic resonance (Arias et al. 2012) and ultrasound (Arias et al. 2013) images. The advantages of the approaches,

such as being able to find any number of surfaces in multiple scans simultaneously without self-intersections, while preserving the topology of the initial surface and enabling point to point correspondences between vertices of the initial mesh and each of the sought surfaces, means that they should be useful for other segmentation tasks as well. Especially applications in which a coarse estimate of the shape of the structure to segment can be found reliably, while the exact boundaries are more difficult to discern. Segmentation of tubular structures, such as blood vessels, from centrelines (Lesage et al. 2009) is one such example. Another is refinement of an initial segmentation found by for example statistical shape (Cootes et al. 1995) and appearance models (Cootes et al. 2001) and atlas registration (Gee et al. 1993). Statistical shape or appearance models are widely used in medical image analysis, however, they can only fit data represented by examples of a training set. The shape and appearance of pathological cases can be very varied and it can be hard, if not impossible, to get examples of all the kinds of pathology one can expect to encounter. Refinement of the statistical model using an optimal surface approach can be used to allow some more flexibility (Sun et al. 2012). In this way the statistical model output can be used as prior knowledge, for instance by using smoothness penalties and constraints to make the final solution follow the shape of the statistical model output to a lesser or greater degree. Point correspondences enable measurements to be compared locally between different surfaces in one image and between surfaces in different images, provided these images can be registered accurately enough. In the case of lung CT images and airway segmentation such accurate registration is only possible with images of the same subject due to the relative large biological differences in lung structures. However, in the case of other modalities and organs, such as for instance brain magnetic resonance images, inter-subject image registration is often accurate enough. An initial segmentation may then be constructed by template atlas registration (Lijn et al. 2012). The refinement of this initial segmentation using a variant of the method presented in Chapter 2 can be used to allow greater flexibility, while still preserving correspondences between atlases registered to images of different subjects, allowing measurements of segmented surfaces to be compared locally between them.

Decreased airway distensibility in COPD could be caused by decreased lung recoil and loss of alveolar attachments due to emphysema (Diaz et al. 2012; Mead et al. 1967), or airway walls could be more rigid because of wall thickening. The longitudinal segmentation method of Chapter 7 is well suited to investigate these reasons. This is especially true if the method can be used to segment ex- and inspiration scan-pairs. Such scan-pairs would allow distensibility to be much more accurately determined, because the scans are at opposite ends of the range of breathing movement within each subject. Expiration scans are, however, particularly difficult to segment, especially in patients with airway disease, where airways can be severely narrowed or entirely collapsed (Ginneken et al. 2008a; Lo et al. 2012). However, the method would

allow these scans to be segmented by using information from the more easy to segment inspiration scan, assuming that they can be registered accurately enough. In general for the method to work, flow lines should start inside, end outside, and only once intersect the airway wall. Flow lines roughly start from the centre of the initial segmentation and so the registration method should register the images accurately enough to put the initial segmentation airway centre inside the actual airway. This puts an upper bound on the amount of registration error the method can tolerate at roughly the lowest airway radii one wishes to segment and analyse. Murphy et al. 2011 compared a range of state-of-the-art registration methods for lung CT images on amongst other things exp- and inspiration scan-pairs. The best performing methods in general achieved landmark errors on these scans of about 1 mm, whereas the average method was closer to 2.5 mm. The best performing methods thus seems to be good enough to allow airways down to and including generation 5 to be analysed in exp- and inspiration images (in the data set of Chapter 6 lumen diameter was measured to be 4.4 ± 1.4 mm in airways of generation 5). Correspondences between points in the segmented exp- and inspiration surfaces could be used to estimate distensibility locally. In this way the method could be used to determine whether lack of distensibility is locally correlated with thicker airway walls or the presence of emphysema, which would provide further insights into the roles of emphysema and small airway disease in causing airflow limitation. Lack of distensibility could be a direct cause of airflow limitation and so the ability to measure this accurately could also lead to new airway signatures of COPD.

List of publications

Papers in international journals

- M. M. W. Wille, L. H. Thomsen, A. Dirksen, J. Petersen, J. H. Pedersen, and S. B. Shaker (submitted). "Emphysema progression is visually detectable in continuous but not in former smokers in a five-year time frame - Interobserver analysis and time-trend of visual assessments of CT scans from the Danish Lung Cancer Screening Trial". In: -.
- A. Feragen*, J. Petersen*, M. Owen, P. Lo, L. H. Thomsen, M. M. W. Wille, A. Dirksen, and M. de Bruijne (submitted). "Geodesic anatomical labeling of airway trees". In: - .
- M. M. W. Wille, J. Petersen, S. B. Shaker, A. Dirksen, L. T. Skovgaard, J. H. Pedersen, and M. de Bruijne (to be submitted). "Airway distensibility, lumen diameter, and wall thickness in COPD - based on CT airway segmentation in the Danish Lung Cancer Screening Trial". In: -.
- J. Petersen, M. M. W. Wille, L. L. Rakêt, A. Feragen, J. H. Pedersen, M. Nielsen, A. Dirksen, and M. de Bruijne (submitted). "Effect of inspiration on airway dimensions measured in maximal inspiration CT images". In: -.
- J. Petersen, M. Nielsen, P. Lo, L. H. Nordenmark, J. H. Pedersen, M. M. W. Wille, A. Dirksen, and M. de Bruijne (2014). "Optimal surface segmentation using flow lines to quantify airway abnormalties in chronic obstructive pulmonary disease". In: *Med Image Anal* (In Press). URL: `http://dx.doi.org/10.1016/j.media.2014.02.004`.

Papers in conference proceedings

- A. Feragen, N. Kasenburg, J. Petersen, M. de Bruijne, and K. Borgwardt (2013b). "Scalable kernels for graphs with continuous attributes". In: *Neural Information Processing Systems (NIPS 2013)*. Vol. 26, pp. 216–224.

*Joint first authors

- J. Petersen, M. Modat, M. J. Cardoso, A. Dirksen, S. Ourselin, and M. de Bruijne (2013c). "Quantitative Airway Analysis in Longitudinal Studies using Groupwise Registration and 4D Optimal Surfaces". In: *Med Image Comput Assist Interv - MICCAI 2013*. Ed. by K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab. Lecture Notes in Computer Science. Springer, pp. 287–294.

- A. Arias, D. Carvalho, J. Petersen, A. van Dijk, A. van der Lugt, W. Niessen, S. Klein, and M. de Bruijne (2013). "Carotid artery lumen segmentation on 3D free-hand ultrasound images using surface graph cuts". In: *Med Image Comput Assist Interv - MICCAI 2013*. Ed. by K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab. Lecture Notes in Computer Science. Springer, pp. 542–549.

- A. Feragen, M. Owen, J. Petersen, M. M. W. Wille, L. H. Thomsen, A. Dirksen, and M. de Bruijne (2013c). "Tree-space statistics and approximations for large-scale analysis of anatomical trees". In: *Information Processing in Medical Imaging*. Ed. by W. Wells, S. Joshi, and K. Pohl. Lecture Notes in Computer Science. Springer, pp. 74–85.

- A. Feragen, J. Petersen, D. Grimm, A. Dirksen, J. H. Pedersen, K. Borgwardt, and M. de Bruijne (2013a). "Geometric tree kernels: Classification of COPD from airway tree geometry". In: *Information Processing in Medical Imaging*. Ed. by W. Wells, S. Joshi, and K. Pohl. Lecture Notes in Computer Science. Springer, pp. 171–183

- A. Feragen, J. Petersen, M. Owen, P. Lo, L. H. Thomsen, M. M. W. Wille, A. Dirksen, and M. de Bruijne (2012). "A Hierarchical Scheme for Geodesic Anatomical Labeling of Airway Trees". In: *Med Image Comput Assist Interv - MICCAI 2012*. Lecture Notes in Computer Science. Springer, pp. 147–155.

- A. Arias, J. Petersen, A. van Engelen, H. Tang, M. Selwaness, J. Witteman, A. van der Lugt, W. Niessen, and M. de Bruijne (2012). "Carotid artery wall segmentation by coupled surface graph cuts". In: *Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging*. Ed. by B. Menze, G. Langs, L. Lu, A. Montillo, Z. Tu, and A. Criminisi, pp. 38–47.

- J. Petersen, V. Gorbunova, M. Nielsen, A. Dirksen, P. Lo, and M. de Bruijne (2011a). "Longitudinal Analysis of Airways using Registration". In: *Fourth International Workshop on Pulmonary Image Analysis*. Ed. by R. Beichel, M. de Bruijne, B. van Ginneken, S. Kabus, A. Kiraly, J. Kuhnigk, J. McClelland, K. Mori, E. van Rikxoort, and S. Rit. CreateSpace, pp. 11–22.

- L. Sørensen, P. Lo, J. Petersen, A. Dirksen, and M. de Bruijne (2011). "Dissimilarity-based classification of anatomical tree structures". In: *Information Processing in Medical Imaging*. Ed. by G. Szkely and H. Hahn. Lecture Notes in Computer Science. Springer.

- J. Petersen, M. Nielsen, P. Lo, Z. Saghir, A. Dirksen, and M. de Bruijne (2011b). "Optimal Graph Based Segmentation Using Flow Lines with Application to Airway Wall Segmentation". In: *Information Processing in Medical Imaging*. Ed. by G. Székely and H. K. Hahn. Lecture Notes in Computer Science. Springer, pp. 49–60.

- J. Petersen, P. Lo, M. Nielsen, G. Edula, H. Ashraf, A. Dirksen, and M. de Bruijne (2010). "Quantitative analysis of airway abnormalities in CT". In: *Medical Imaging: Computer-Aided Diagnosis*. Ed. by N. Karssemeijer and R. Summers. Proceedings of SPIE 7624. SPIE Press.

Published abstracts

- M. M. Wille, J. Petersen, A. Dirksen, and M. de Bruijne (2013). "Airway distensibility in chronic obstructive pulmonary disease - evaluation by CT airway segmentation and lung density measurement based on the Danish Lung Cancer Screening Trial". In: *American Thoracic Society International Conference (ATS)*.

- J. Petersen, A. Feragen, M. Owen, M. M. Wille, L. H. Thomsen, A. Dirksen, and M. de Bruijne (2013a). "Automatic system for segmentation, longitudinal matching, anatomical labeling and measurements of airways from CT images". In: *Novel technology that shapes Radiology: EIBIR presents IMAGINE*.

- J. Petersen, A. Feragen, L. H. Thomsen, M. M. Wille, A. Dirksen, and M. de Bruijne (2013b). "Manual airway labeling has limited reproducibility". In: *ECR - Annual meeting of the European Society of Radiology*.

- J. Petersen, M. M. Wille, L. H. Thomsen, A. Feragen, A. Dirksen, and M. de Bruijne (2013d). "The effect of inspiration on airway dimensions measured in CT images from the Danish Lung Cancer Screening Trial". In: *ECR - Annual meeting of the European Society of Radiology*.

- M. M. Wille, L. H. Thomsen, J. Petersen, S. B. Shaker, A. Dirksen, and J. H. Pedersen (2012). "Interobserver variability in visual evaluation of thoracic CT scans and comparison with automatic computer measurements of CT lung density". In: *European Respiratory Society, Annual Congres, Vienna*.

Bibliography

Abràmoff, M. D., R. Kardon, S. R. Russel, X. Wu, and M. Sonka (2008). "Intraretinal Layer Segmentation of Macular Optical Coherence Tomography Images Using Optimal 3-D Graph Search". In: *IEEE Trans Med Imag* 27.10, pp. 1495–1505.

Achenbach, T., O. Weinheimer, A. Biedermann, S. Schmitt, D. Freudenstein, E. Goutham, R. P. Kunz, R. Buhl, C. Dueber, and C. P. Heussel (2008). "MDCT assessment of airway wall thickness in COPD patients using a new method: correlations with pulmonary function tests". In: *Eur Radiol* 18, pp. 2731–2738.

Arias, A., J. Petersen, A. van Engelen, H. Tang, M. Selwaness, J. Witteman, A. van der Lugt, W. Niessen, and M. de Bruijne (2012). "Carotid artery wall segmentation by coupled surface graph cuts". In: *Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging*. Ed. by B. Menze, G. Langs, L. Lu, A. Montillo, Z. Tu, and A. Criminisi, pp. 38–47.

Arias, A., D. Carvalho, J. Petersen, A. van Dijk, A. van der Lugt, W. Niessen, S. Klein, and M. de Bruijne (2013). "Carotid artery lumen segmentation on 3D free-hand ultrasound images using surface graph cuts". In: *Med Image Comput Assist Interv - MICCAI 2013*. Ed. by K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab. Lecture Notes in Computer Science. Springer, pp. 542–549.

Ashraf, H., P. Lo, S. B. Shaker, M. de Bruijne, A. Dirksen, P. Tønnesen, M. Dahlbäck, and J. H. Pedersen (2011). "Short-term effect of changes in smoking behaviour on emphysema quantification by CT". In: *Thorax* 66.1, pp. 55–60.

Baldi, S., R. Dellaca, L. Govoni, R. Torchio, A. Aliverti, P. Pompilio, L. Corda, C. Tantucci, C. Gulotta, V. Brusasco, and R. Pellegrino (2010). "Airway distensibility and volume recruitment with lung inflation in COPD". In: *J Appl Physiol* 109.4, pp. 1019–26.

Bates, D., M. Maechler, and B. Bolker (2012). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999999-0. URL: `http://CRAN.R-project.org/package=lme4`.

Berger, P., V. Perot, P. Desbarats, J. M. T. de Lara, R. Marthan, and F. Laurent (2005). "Airway Wall Thickness in Cigarette Smokers: Quantitative Thin-Section CT assessment". In: *Radiology* 235, pp. 1055–1064.

Billera, L. J., S. P. Holmes, and K. Vogtmann (2001). "Geometry of the space of phylogenetic trees". In: *Adv in Appl Math* 27.4, pp. 733–767.

Bogunovic, H., J. Pozo, R. Cardenes, L. San Roman, and A. Frangi (2013). "Anatomical Labeling of the Circle of Willis using Maximum A Posteriori Probability Estimation". In: *IEEE Transactions on Medical Imaging* 32.9, pp. 1587–99.

Boissonnat, J.-D. and S. Oudot (2005). "Provably good sampling and meshing of surfaces". In: *Graphical Models* 67.5, pp. 405–451.

Boykov, Y. and V. Kolmogorov (2004). "An Experimental Comparison of Min-cut/Max-Flow Algorithms for Energy Minimization in Vision". In: *IEEE Trans Pattern Anal Machine Intell* 26.9, pp. 1124–1137.

Brown, R. H., C. Herold, E. A. Zerhouni, and W. Mitzner (1994). "Spontaneous airways constrict during breath holding studied by high-resolution computed tomography". In: *Chest* 106.3, pp. 920–4.

Brown, R. H., N. Scichilone, B. Mudge, F. B. Diemer, S. Permutt, and A. Togias (2001). "High-resolution computed tomographic evaluation of airway distensibility and the effects of lung inflation on airway caliber in healthy subjects and individuals with asthma". In: *Am J Respir Crit Care Med* 163.4, pp. 994–1001.

Bülow, T., C. Lorenz, R. Wiemker, and J. Honko (2006). "Point based methods for automatic bronchial tree matching and labeling". In: *SPIE Medical Imaging*. Vol. 6143, pp. 225–234.

Cardoso, M. J., M. J. Clarkson, G. R. Ridgway, M. Modat, N. C. Fox, and S. Ourselin (2011). "LoAd: A Locally Adaptive Cortical Segmentation Algorithm". In: *NeuroImage* 56.3, pp. 1386–1397.

Chmura, K., S. Hines, and E. D. Chan (2008). *CT of the Airways*. Ed. by P. Boiselle and D. Lynch. Contemporary medical imaging. Humana Press, pp. 3–24. ISBN: 9781597451390. URL: `http://books.google.dk/books?id=QaoKfnl-qIcC`.

Ciba 1959 (1959). "Ciba Guest Symposium Report: Definitions and Classifications of Chronic Obstructive Pulmonary Emphysema and Related Conditions". In: *Thorax* 14, pp. 286–299.

Cootes, T. F., C. J. Taylor, D. H. Cooper, and J. Graham (1995). "Active shape models - their training and application". In: *Computer Vision and Image Understanding* 61.1, pp. 38–59.

Cootes, T. F., G. J. Edwards, and C. J. Taylor (2001). "Active Appearance Models". In: *IEEE Trans Pattern Anal Machine Intell* 23.6, pp. 681–685.

Cosio, M., H. Ghezzo, J. C. Hogg, R. Corbin, M. Loveland, J. Dosman, and P. T. Macklem (1978). "The relations between structural changes in small airways and pulmonary-function tests". In: *N Engl J Med* 298.23, pp. 1277–81.

Cosio, M. G. and A. Guerassimov (1999). "Chronic obstructive pulmonary disease. Inflammation of small airways and lung parenchyma". In: *Am J Respir Crit Care Med* 160.5 Pt 2, S21–S25.

Decramer, M., W. Janssens, and M. Miravitlles (2012). "Chronic Obstructive Pulmonary Disease". In: *Lancet* 379, pp. 1341–51.

Dey, T. K. and J. Sun (2006). "Normal and Feature Estimation from Noisy Point Clouds". In: *Proceedings of the 26th International Conference on Foundations of Software Technology and Theoretical Computer Science*, pp. 21–32.

Diaz, A., C. Valim, T. Yamashiro, R. Estépar, J. Ross, S. Matsuoka, B. Bartholmai, H. Hatabu, E. Silverman, and G. Washko (2010a). "Airway count and emphysema assessed by chest CT imaging predicts clinical outcome in smokers". In: *Chest* 138.4, pp. 880–887.

Diaz, A., C. Come, J. Ross, R. San José Estépar, M. Han, S. Loring, E. Silverman, and G. Washko (2012). "Association between airway caliber changes with lung inflation and emphysema assessed by volumetric CT scan in subjects with COPD". In: *Chest* 141.3, pp. 736–44.

Diaz, A. A., B. Bartholmai, R. S. J. Estépar, J. Ross, S. Matsuoka, T. Yamashiro, H. Hatabu, J. J. Reilly, E. K. Silverman, and G. R. Washko (2010b). "Relationship of Emphysema and Airway Disease assessed by CT to Exercise Capacity in COPD". In: *Respir Med* 104.8, pp. 1145–1151.

Dijkstra, A. E., D. S. Postma, N. ten Hacken, J. M. Vonk, M. Oudkerk, P. M. van Ooijen, P. Zanen, F. A. M. Hoesein, B. van Ginneken, M. Schmidt, and H. J. Groen (2013). "Low-dose CT measurements of airway dimensions and emphysema associated with airflow limitation in heavy smokers: a cross sectional study". In: *Respir Res* 14.11.

Dirksen, A., J. H. Dijkman, F. Madsen, B. Stoel, D. C. Hutchison, C. S. Ulrik, L. T. Skovgaard, A. Kok-Jensen, A. Rudolphus, N. Seersholm, H. A. Vrooman, J. H. Reiber, N. C. Hansen, T. Heckscher, K. Viskum, and J. Stolk (1999). "A randomized clinical trial of alpha(1)-antitrypsin augmentation therapy". In: *Am J Respir Crit Care Med* 160.5 Pt 1, pp. 1468–72.

Dirksen, A. (2008). "Monitoring the progress of emphysema by repeat computed tomography scans with focus on noise reduction". In: *Proc Am Thorac Soc* 5.9, pp. 925–8.

Estépar, R. S., G. G. Washko, E. K. Silverman, J. J. Reilly, R. Kikinis, and C. F. Westin (2006). "Accurate airway wall estimation using phase congruency". In: *Med Image Comput Assist Interv - MICCAI 2006*. Vol. 9. Lecture Notes in Computer Science Pt 2, pp. 125–34.

Fehlberg, E. (1970). "Klassische Runge-Kutta-Formeln vierter und niedrigerer Ordnung mit Schrittweiten-Kontrolle und ihre Anwendung auf Wärmeleitungsprobleme". In: *Computing (Arch Elektron Rechnen)* 6, pp. 61–71.

Feneis, H. (1995). *Anatomisk billedordbog*. Forlaget Munksgaard, Copenhagen.

Feragen, A., P. Lo, V. Gorbunova, M. Nielsen, A. Dirksen, J. Reinhardt, F. Lauze, and M. de Bruijne (2011). "An airway tree-shape model for geodesic airway branch labeling". In: *MICCAI WS on Math Found Comp Anat.*

Feragen, A., J. Petersen, M. Owen, P. Lo, L. H. Thomsen, M. M. W. Wille, A. Dirksen, and M. de Bruijne (2012). "A Hierarchical Scheme for Geodesic Anatomical Labeling of Airway Trees". In: *Med Image Comput Assist Interv - MICCAI 2012*. Lecture Notes in Computer Science. Springer, pp. 147–155.

Feragen, A., J. Petersen, D. Grimm, A. Dirksen, J. H. Pedersen, K. Borgwardt, and M. de Bruijne (2013a). "Geometric tree kernels: Classification of COPD from airway tree geometry". In: *Information Processing in Medical Imaging*. Ed. by W. Wells, S. Joshi, and K. Pohl. Lecture Notes in Computer Science. Springer, pp. 171–183.

Feragen, A., N. Kasenburg, J. Petersen, M. de Bruijne, and K. Borgwardt (2013b). "Scalable kernels for graphs with continuous attributes". In: *Neural Information Processing Systems (NIPS 2013)*. Vol. 26, pp. 216–224.

Feragen, A., M. Owen, J. Petersen, M. M. W. Wille, L. H. Thomsen, A. Dirksen, and M. de Bruijne (2013c). "Tree-space statistics and approximations for large-scale analysis of anatomical trees". In: *Information Processing in Medical Imaging*. Ed. by W. Wells, S. Joshi, and K. Pohl. Lecture Notes in Computer Science. Springer, pp. 74–85.

Freimanis, A. K. and W. Molnar (1960). "Chronic bronchitis and emphysema at bronchography: survey of diagnostic features obtained by reviewing 2,000 bronchograms". In: *Radiology* 74, pp. 194–205.

Fujimoto, K., Y. Kitaguchi, K. Kubo, and T. Honda (2006). "Clinical analysis of chronic obstructive pulmonary disease phenotypes classified using high-resolution computed tomography". In: *Respirology* 11.6, pp. 731–40.

Galassi, M., J. Davies, J. Theiler, B. Gough, G. Jungman, P. Alken, M. Booth, and F. Rossi. *GNU Scientific Library Reference Manual (3rd Ed.)* ISBN: 0954612078. URL: `http://www.gnu.org/software/gsl/`.

Gee, J., M. Reivich, and R. Bajcsy (1993). "Elastically Deforming a Three-Dimensional Atlas to Match Anatomical Brain Images". In: *Journal of Computer Assisted Tomography* 17.1, pp. 225–236.

Georgopoulos, D., E. Giannouli, and D. Patakas (1993). "Effects of extrinsic positive end-expiratory pressure on mechanically ventilated patients with chronic obstructive pulmonary disease and dynamic hyperinflation". In: *Intensive Care Med* 19.4, pp. 197–203.

Gierada, D. S., R. D. Yusen, T. K. Pilgram, L. Crouch, R. M. Slone, K. T. Bae, S. S. Lefrak, and J. D. C. JD (2001). "Repeatability of quantitative CT indexes of emphysema in patients evaluated for lung volume reduction surgery". In: *Radiology* 220.2, pp. 448–54.

Ginneken, B. van, W. Baggerman, and E. van Rikxoort (2008a). "Robust Segmentation and Anatomical Labeling of the Airway Tree from Thoracic CT Scans". In: *Med Image Comput Assist Interv - MICCAI 2008*. Lec-

ture Notes in Computer Science Pt II. Berlin, Heidelberg: Springer-Verlag, pp. 219–226.

Ginneken, B. van, W. Baggerman, and E. M. van Rikxoort (2008b). "Robust Segmentation and Anatomical Labeling of the Airway Tree from Thoracic CT Scans". In: *Med Image Comput Assist Interv - MICCAI 2008*. Lecture Notes in Computer Science, pp. 219–226.

GOLD 2013. *Global Strategy for the Diagnosis, Management and Prevention of COPD*. Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2013. Available from: http://www.goldcopd.org/.

Gorbunova, V., J. Sporring, M. Nielsen, A. Dirksen, P. Lo, and M. de Bruijne (2012). "Mass preserving image registration for lung CT". In: *Med Image Anal* 16.4, pp. 786–95.

Gorbunova, V., P. Lo, H. Ashraf, A. Dirksen, M. Nielsen, and M. de Bruijne (2008). "Weight Preserving Image Registration for Monitoring Disease Progression in Lung CT". In: *Med Image Comput Assist Interv - MICCAI 2008*. Lecture Notes in Computer Science Pt II. Berlin, Heidelberg: Springer-Verlag, pp. 863–870.

Gould, G. A., W. MacNee, A. McLean, P. M. Warren, A. Redpath, J. J. Best, D. Lamb, and D. C. Flenley (1988). "CT measurements of lung density in life can quantitate distal airspace enlargement–an essential defining feature of human emphysema". In: *Am Rev Respir Dis* 137.2, pp. 380–92.

Graham, M. W. and W. E. Higgins (2006). "Optimal graph-theoretic approach to 3D anatomical tree matching". In: *International Symposium on Biomedical Imaging (ISBI)*, pp. 109–112.

Grydeland, T. B., A. Dirksen, H. O. Coxson, T. M. Eagan, E. Thorsen, S. G. Pillai, S. Sharma, G. E. Eide, A. Gulsvik, and P. S. Bakke (2010). "Quantitative computed tomography measures of emphysema and airway wall thickness are related to respiratory symptoms". In: *Am J Respir Crit Care Med* 181.4, pp. 353–9.

Gu, S., Z. Wang, J. M. Siegfried, D. Wilson, W. L. Bigbee, and J. Pu (Jan. 2012). "Automated Lobe-based Airway Labeling". In: *Journal of Biomedical Imaging* 2012.

Hackx, M., A. A. Bankier, and P. A. Gevenois (2012). "Chronic Obstructive Pulmonary Disease: CT Quantification of Airways Disease". In: *Radiology* 265.1, pp. 34–48.

Haeker, M., X. Wu, M. Abrámoff, R. Kardon, and M. Sonka (2007). "Incorporation of Regional Information in Optimal 3-D Graph Search with Application for Intraretinal Layer Segmentation of Optical Coherence Tomography Images". In: *Information Processing in Medical Imaging*. Ed. by N. Karssemeijer and B. Lelieveldt. Vol. 4584. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 607–618.

Hasegawa, M., Y. Nasuhara, Y. Onodera, H. Makita, K. Nagai, S. Fuke, Y. Ito, T. Betsuyaku, and M. Nishimura (2006). "Airflow limitation and airway

dimensions in chronic obstructive pulmonary disease." In: *Am J Respir Crit Care Med* 173.12, pp. 1309–15.

Hogg, J. C., P. T. Macklem, and W. M. Thurlbeck (1968). "Site and nature of airway obstruction in chronic obstructive lung disease". In: *N Engl J Med* 278.25, pp. 1355–1360.

Hogg, J. C., S. J. Nepszy, P. T. Macklem, and W. M. Thurlbeck (1969). "Elastic properties of the centrilobular emphysematous space". In: *J Clin Invest* 48.7, pp. 1306–1312.

Hogg, J. C., F. Chu, S. Utokaparch, R. Woods, W. M. Elliot, L. Buzatu, R. M. Cherniack, R. M. Rogers, F. C. Sciurba, H. O. Coxson, and P. D. Paré (2004). "The Nature of Small-Airway Obstruction in Chronic Obstructive Pulmonary Disease". In: *N Engl J Med* 350.26, pp. 2645–2653.

Ishikawa, H. (2003). "Exact Optimization for Markov Random Fields with Convex Priors". In: *IEEE Trans Pattern Anal Machine Intell* 25.10, pp. 1333–1336.

Johannessen, A., T. D. Skorge, M. Bottai, T. B. Grydeland, R. M. Nilsen, H. Coxson, A. Dirksen, E. Omenaas, A. Gulsvik, and P. Bakke (2013). "Mortality by level of emphysema and airway wall thickness". In: *Am J Respir Crit Care Med* 187.6, pp. 602–8.

Jones, P. W. (2009). "Health Status and the Spiral of Decline". In: *COPD: Journal of Chronic Obstructive Pulmonary Disease* 6.1, pp. 59–63.

Jong, P. A. de, N. L. Müller, P. D. Paré, and H. O. Coxson (2005). "Computed tomographic imaging of the airways: relationship to structure and function". In: *Eur Respir J* 26.1, pp. 140–52.

Jong, P. A. de, F. R. Long, and Y. Nakano (2006). "Computed tomography dose and variability of airway dimension measurements: how low can we go?" In: *Pediatr Radiol* 36.10, pp. 1043–7.

Kaftan, J., A. Kiraly, D. Naidich, and C. Novak (2006). "A novel multi-purpose tree and path matching algorithm with application to airway trees". In: *SPIE Medical Imaging*. Vol. 6143.

Kainmueller, D., H. Lamecker, M. O. Heller, B. Weber, H.-C. Hege, and S. Zachow (2013). "Omnidirectional displacements for deformable surfaces". In: *Med Image Anal* 17, pp. 429–441.

King, G. G., N. L. Müller, K. P. Whittall, Q. S. Xiang, and P. D. Paré (2000). "An analysis algorithm for measuring airway lumen and wall areas from high-resolution computed tomography data". In: *Am J Respir Crit Care Med* 161(2 Pt 1), pp. 574–80.

Kitaguchi, Y., K. Fujimoto, K. Kubo, and T. Honda (2006). "Characteristics of COPD phenotypes classified according to the findings of HRCT". In: *Respir Med* 100.10, pp. 1742–52.

Kitaoka, H., Y. Park, J. Tschirren, J. M. Reinhardt, M. Sonka, G. McLennan, and E. A. Hoffman (2002). "Automated Nomenclature Labeling of the Bronchial Tree in 3D-CT Lung Images". In: *Med Image Comput Assist Interv - MICCAI 2002*. Lecture Notes in Computer Science, pp. 1–11.

Lederlin, M., F. Laurent, Y. Portron, A. Ozier, H. Cochet, P. Berger, and M. Montaudon (2012). "CT attenuation of the bronchial wall in patients with asthma: comparison with geometric parameters and correlation with function and histologic characteristics". In: *AJR Am J Roentgenol* 199.6, pp. 1226–33.

Lesage, D., E. D. Angelini, I. Bloch, and G. Funka-Lea (2009). "A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes". In: *Med Image Anal* 13.6, pp. 819–845.

Li, K., X. Wu, D. Z. Chen, and M. Sonka (2006). "Optimal Surface Segmentation in Volumetric Images - A Graph-Theoretic Approach". In: *IEEE Trans Pattern Anal Machine Intell* 28.1, pp. 119–134.

Lijn, F. van der, M. de Bruijne, S. Klein, T. den Heijer, Y. Y. Hoogendam, A. van der Lugt, M. M. B. Breteler, and W. J. Niessen (2012). "Automated Brain Structure Segmentation Based on Atlas Registration and Appearance Models". In: *IEEE Trans Med Imag* 31.2, pp. 276–286.

Liu, X., D. Z. Chen, M. Tawhai, X. Wu, E. Hoffman, and M. Sonka (2012). "Optimal graph search based segmentation of airway tree double surfaces across bifurcations". In: *IEEE Trans Med Imag* (99).

Lo, P. (2010). "Segmentation of Lung Structures in CT". PhD. Department of Computer Science, University of Copenhagen, Denmark.

Lo, P., J. Sporring, H. Ashraf, J. J. Pedersen, and M. de Bruijne (2010). "Vessel-guided airway tree segmentation: A voxel classification approach". In: *Med Image Anal* 14.4, pp. 527–38.

Lo, P., E. van Rikxoort, J. Goldin, F. Abtin, M. de Bruijne, and M. Brown (2011). "A Bottom-up approach for labeling of human airway trees". In: *Fourth International Workshop on Pulmonary Image Analysis*. Ed. by R. Beichel, M. de Bruijne, B. van Ginneken, S. Kabus, A. Kiraly, J. Kuhnigk, J. McClelland, K. Mori, E. van Rikxoort, and S. Rit. CreateSpace.

Lo, P., J. Sporring, H. Ashraf, J. J. H. Pedersen, and M. de Bruijne (2008). "Vessel-guided airway segmentation based on voxel classification". In: *The first International Workshop on Pulmonary Image Analysis* 1, pp. 113–122.

Lo, P., J. Sporring, J. J. H. Pedersen, and M. de Bruijne (2009). "Airway Tree Extraction with Locally Optimal Paths". In: *Med Image Comput Assist Interv - MICCAI 2009*. Ed. by G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, and C. Taylor. Vol. 5762. Lecture Notes in Computer Science. Springer, pp. 51–58.

Lo, P., B. van Ginneken, J. M. Reinhardt, T. Yavarna, P. A. de Jong, B. Irving, C. Fetita, M. Ortner, R. Pinho, J. Sijbers, M. Feuerstein, A. Fabijanska, C. Bauer, R. Beichel, C. S. Mendoza, R. Wiemker, J. Lee, A. P. Reeves, S. Born, O. Weinheimer, E. M. van Rikxoort, J. Tschirren, K. Mori, B. Odry, D. P. Naidich, I. Hartmann, E. A. Hoffman, M. Prokop, J. H. Pedersen, and M. de Bruijne (2012). "Extraction of Airways from CT (EXACT'09)". In: *IEEE Trans Med Imag* 31 (11), pp. 2093–2107.

Lopez, A. D., K. Shibuya, C. Rao, C. D. Mathers, A. L. Hansell, L. S. Held, V. Schmid, and S. Buist (2006). "Chronic obstructive pulmonary disease: current burden and future projections". In: *Eur Respir J* 27.2, pp. 397–412.

Lutey, B. A., S. H. Conradi, J. J. Atkinson, J. Zheng, K. B. Schechtman, R. M. Senior, and D. S. Gierada (2013). "Accurate measurement of small airways on low-dose thoracic CT scans in smokers". In: *Chest* 143.5, pp. 1321–9.

Martinez, F. J., G. Foster, J. L. Curtis, G. Criner, G. Weinmann, A. Fishman, M. M. DeCamp, J. Benditt, F. Sciurba, B. Make, Z. Mohsenifar, P. Diaz, E. Hoffman, R. Wise, and NETT Research Group (2006). "Predictors of mortality in patients with emphysema and severe airflow obstruction". In: *Am J Respir Crit Care Med* 173.12, pp. 1326–34.

Matsuoka, S., Y. Kurihaha, K. Yagihashi, M. Hoshino, and Y. Nakajima (2008). "Airway Dimensions at Inspiratory and Expiratory Multisection CT in Chronic Obstructive Pulmonary Disease: Correlation with Airflow Limitation". In: *Radiology* 248.3, pp. 1042–9.

McDonough, J. E., R. Yuan, M. Suzuki, N. Seyednejad, W. M. Elliott, P. G. Sanchez, A. C. Wright, W. B. Gefter, L. Litzky, H. O. Coxson, P. D. Paré, D. D. Sin, R. A. Pierce, J. C. Woods, A. M. McWilliams, J. R. Mayo, S. C. Lam, J. D. Cooper, and J. C. Hogg (2011). "Small-Airway Obstruction and Emphysema in Chronic Obstructive Pulmonary Disease". In: *N Engl J Med* 365, pp. 1567–75.

Mead, J., J. M. Turner, P. T. Macklem, and J. B. Little (1967). "Significance of the relationship between lung recoil and maximum expiratory flow". In: *J Appl Physiol* 22.1, pp. 95–108.

Metzen, J. H., T. Kröger, A. Schenk, S. Zidowitz, H.-O. Peitgen, and X. Jiang (2009). "Matching of Anatomical Tree Structures for Registration of Medical Images". In: *Image and Vision Computing* 27, pp. 923–933.

Miller, M. R., R. Crapo, J. Hankinson, V. Brusasco, F. Burgos, R. Casaburi, A. Coates, P. Enright, C. P. van der Grinten, P. Gustafsson, R. Jensen, D. C. Johnson, N. MacIntyre, R. McKay, D. Navajas, O. F. Pedersen, R. Pellegrino, G. Viegi, J. Wanger, and A. T. Force" (2005). "General considerations for lung function testing". In: *Eur Respir J.* 26.1, pp. 153–61.

Modat, M., G. R. Ridgway, Z. A. Taylor, M. Lehmann, J. Barnes, D. J. Hawkes, N. C. Fox, and S. Ourselin (2010). "Fast free-form deformation using graphics processing units". In: *Comput Methods Programs Biomed* 98.3, pp. 278–84.

Moon, W. K., J. G. Im, K. M. Yeon, and M. C. Han (1997). "Tuberculosis of the central airways: CT findings of active and fibrotic disease". In: *AJR Am J Roentgenol* 169.3, pp. 649–53.

Mori, K., S. Ota, D. Deguchi, T. Kitasaka, Y. Suenaga, S. Iwano, Y. Hasegawa, H. Takabatake, M. Mori, and H. Natori (2009). "Automated Anatomical Labeling of Bronchial Branches Extracted from CT Datasets Based on

Machine Learning and Combination Optimization and Its Application to Bronchoscope Guidance". In: *Med Image Comput Assist Interv - MICCAI 2009*. Vol. 5762. Lecture Notes in Computer Science, pp. 707–714.

Müller, N. L., C. A. Staples, R. R. Miller, and R. T. Abboud (1988). ""Density mask". An objective method to quantitate emphysema using computed tomography". In: *Chest* 94.4, 782–787.

Murphy, K., B. van Ginneken, J. M. Reinhardt, S. Kabus, K. Ding, X. Deng, K. Cao, K. Du, G. E. Christensen, V. Garcia, T. Vercauteren, N. Ayache, O. Commowick, G. Malandain, B. Glocker, N. Paragios, N. Navab, V. Gorbunova, J. Sporring, M. de Bruijne, X. Han, M. P. Heinrich, J. A. Schnabel, M. Jenkinson, C. Lorenz, M. Modat, J. R. McClelland, S. Ourselin, S. E. Muenzing, M. A. Viergever, D. D. Nigris, D. L. Collins, T. Arbel, M. Peroni, R. Li, G. C. Sharp, A. Schmidt-Richberg, J. Ehrhardt, R. Werner, D. Smeets, D. Loeckx, G. Song, N. Tustison, B. Avants, J. C. Gee, M. Staring, S. Klein, B. C. Stoel, M. Urschler, M. Werlberger, J. Vandemeulebroucke, S. Rit, D. Sarrut, and J. P. Pluim (2011). "Evaluation of Registration Methods on Thoracic CT: The EMPIRE10 Challenge". In: *IEEE Trans Med Imag* 30.11, pp. 1920–20.

Nakano, Y., S. Muro, H. Sakai, T. Hirai, K. Chin, M. Tsukino, K. Nishimura, H. Itoh, P. D. Paré, J. C. Hogg, and M. Mishima (2000). "Computed Tomographic Measurements of Airway Dimensions and Emphysema in Smokers - Correlation with Lung Function". In: *Am J Respir Crit Care Med* 162, pp. 1102–1108.

Nakano, Y., J. C. Wong, P. A. de Jong, L. Buzatu, T. Nagao, H. O. Coxson, W. M. Elliott, J. C. Hogg, and P. D. Paré (2005). "The Prediction of Small Airway Dimensions Using Computed Tomography". In: *Am J Respir Crit Care Med* 171, pp. 142–146.

Netter, F. (1989). *Atlas of human anatomy*. Novartis: East Hanover: NJ.

O'Donnell, D. E., R. Sanii, N. R. Anthonisen, and M. Younes (1987). "Effect of dynamic airway compression on breathing pattern and respiratory sensation in severe chronic obstructive pulmonary disease". In: *Am Rev Respir Dis* 135.4, pp. 912–8.

Ohara, T., T. Hirai, S. Sato, K. Terada, D. Kinose, A. Haruna, S. Marumo, M. Nishioka, E. Ogawa, Y. Nakano, Y. Hoshino, Y. Ito, H. Matsumoto, A. Niimi, T. Mio, K. Chin, S. Muro, and M. Mishima (2008). "Longitudinal study of airway dimensions in chronic obstructive pulmonary disease using computed tomography." In: *Respirology* 13.3, pp. 372–8.

Ortner, M., C. Fetita, P.-Y. Brillet, F. Prêteux, and P. Grenier (2010). "3D Vector Flow Guided Segmentation of Airway Wall in MSCT". In: *Advances in Visual Computing*. Ed. by G. Bebis, R. D. Boyle, B. Parvin, D. Koracin, R. Chung, R. I. Hammoud, M. Hussain, K.-H. Tan, R. Crawfis, D. Thalmann, D. Kao, and L. Avila. Lecture Notes in Computer Science 2. Springer, pp. 302–311.

Owen, M. and J. S. Provan (2011). "A Fast Algorithm for Computing Geodesic Distances in Tree Space". In: *ACM/IEEE Transactions on Computational Biology and Bioinformatics* 8, pp. 2–13.

Pedersen, J. H., H. Ashraf, A. Dirksen, K. Bach, H. Hansen, P. Toennesen, H. Thorsen, J. Brodersen, B. G. Skov, M. Døssing, J. Mortensen, K. Richter, P. Clementsen, and N. Seersholm (2009). "The Danish randomized lung cancer CT screening trial– overall design and results of the prevalence round." In: *J Thorac Oncol* 4.5, pp. 608–614.

Pellegrino, R., G. Viegi, V. Brusasco, R. O. Crapo, F. Burgos, R. Casaburi, A. Coates, C. P. van der Grinten, P. Gustafsson, J. Hankinson, R. Jensen, D. C. Johnson, N. MacIntyre, R. McKay, M. R. Miller, D. Navajas, O. F. Pedersen, and J. Wanger (2005). "Interpretative strategies for lung function tests". In: *Eur Respir J* 26.5, pp. 948–68.

Petersen, J., P. Lo, M. Nielsen, G. Edula, H. Ashraf, A. Dirksen, and M. de Bruijne (2010). "Quantitative analysis of airway abnormalities in CT". In: *Medical Imaging: Computer-Aided Diagnosis*. Ed. by N. Karssemeijer and R. Summers. Proceedings of SPIE 7624. SPIE Press.

Petersen, J., V. Gorbunova, M. Nielsen, A. Dirksen, P. Lo, and M. de Bruijne (2011a). "Longitudinal Analysis of Airways using Registration". In: *Fourth International Workshop on Pulmonary Image Analysis*. Ed. by R. Beichel, M. de Bruijne, B. van Ginneken, S. Kabus, A. Kiraly, J. Kuhnigk, J. McClelland, K. Mori, E. van Rikxoort, and S. Rit. CreateSpace, pp. 11–22.

Petersen, J., M. Nielsen, P. Lo, Z. Saghir, A. Dirksen, and M. de Bruijne (2011b). "Optimal Graph Based Segmentation Using Flow Lines with Application to Airway Wall Segmentation". In: *Information Processing in Medical Imaging*. Ed. by G. Székely and H. K. Hahn. Lecture Notes in Computer Science. Springer, pp. 49–60.

Petersen, J., A. Feragen, M. Owen, M. M. Wille, L. H. Thomsen, A. Dirksen, and M. de Bruijne (2013a). "Automatic system for segmentation, longitudinal matching, anatomical labeling and measurements of airways from CT images". In: *Novel technology that shapes Radiology: EIBIR presents IMAGINE*.

Petersen, J., A. Feragen, L. H. Thomsen, M. M. Wille, A. Dirksen, and M. de Bruijne (2013b). "Manual airway labeling has limited reproducibility". In: *ECR - Annual meeting of the European Society of Radiology*.

Petersen, J., M. Modat, M. J. Cardoso, A. Dirksen, S. Ourselin, and M. de Bruijne (2013c). "Quantitative Airway Analysis in Longitudinal Studies using Groupwise Registration and 4D Optimal Surfaces". In: *Med Image Comput Assist Interv - MICCAI 2013*. Ed. by K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab. Lecture Notes in Computer Science. Springer, pp. 287–294.

Petersen, J., M. M. Wille, L. H. Thomsen, A. Feragen, A. Dirksen, and M. de Bruijne (2013d). "The effect of inspiration on airway dimensions measured

in CT images from the Danish Lung Cancer Screening Trial". In: *ECR - Annual meeting of the European Society of Radiology.*

Petersen, J., M. Nielsen, P. Lo, L. H. Nordenmark, J. H. Pedersen, M. M. W. Wille, A. Dirksen, and M. de Bruijne (2014). "Optimal surface segmentation using flow lines to quantify airway abnormalties in chronic obstructive pulmonary disease". In: *Med Image Anal* (In Press). URL: http://dx.doi.org/10.1016/j.media.2014.02.004.

Pisupati, C., L. Wolff, W. Mitzner, and E. Zerhouni (1996). "Tracking 3D pulmonary tree structures". In: *Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis*, pp. 160–169. DOI: 10.1109/MMBIA.1996.534068.

Pu, J., J. K. Leader, X. Meng, B. Whiting, D. Wilson, F. C. Sciurba, J. J. Reilly, W. L. Bigbee, J. Siegfried, and D. Gur (2012). "Three-dimensional airway tree architecture and pulmonary function". In: *Acad Radiol* 19.11, pp. 1395–401.

Quanjer, P. H., G. J. Tammeling, J. E. Cotes, O. F. Pedersen, R. Peslin, and J. C. Yernault (1993). "Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society". In: *Eur Respir J Suppl* 16, pp. 5–40.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org/.

Robinson, D. and L. Foulds (1979). "Comparison of weighted labeled trees". In: *Combinatorial Mathematics VI*. Vol. 748. Lecture Notes in Mathematics, pp. 119–126.

Rutten, E. P. A., T. B. Grydeland, S. G. Pillai, S. Wagers, A. Dirksen, H. O. Coxson, A. Gulsvik, E. F. M. Wouters, and P. S. Bakke (2011). "Quantitative CT: Associations between Emphysema, Airway Wall Thickness and Body Composition in COPD". In: *Pulm Med* 419328.

Saragaglia, A., C. Fetita, and F. Prêteux (2006). "Assessment of Airway Remodeling in Asthma: Volumetric Versus Surface Quantification Approaches". In: *Med Image Comput Assist Interv - MICCAI 2006*. Vol. 9. Lecture Notes in Computer Science Pt 2, pp. 413–20.

Scichilone, N., T. Kapsali, S. Permutt, and A. Togias (2000). "Deep inspiration-induced bronchoprotection is stronger than bronchodilation". In: *Am J Respir Crit Care Med* 162.3 Pt 1, pp. 910–6.

Scichilone, N., S. Permutt, and A. Togias (2001). "The lack of the bronchoprotective and not the bronchodilatory ability of deep inspiration is associated with airway hyperresponsiveness". In: *Am J Respir Crit Care Med* 163.2, pp. 413–9.

Shaker, S. B., A. Dirksen, P. Lo, L. T. Skovgaard, M. de Bruijne, and J. H. Pedersen (2012). "Factors influencing the decline in lung density in a

Danish lung cancer screening cohort". In: *Eur Respir J* 40.5, pp. 1142–1148.

Sieren, J. P., J. J. D. Newell, P. F. Judy, D. A. Lynch, K. S. Chan, J. Guo, and E. A. Hoffman (2012). "Reference standard and statistical model for intersite and temporal comparisons of CT attenuation in a multicenter quantitative lung study". In: *Med Phys* 39.9.

Simon, G. and H. J. Galbraith (1953). "Radiology of chronic bronchitis". In: *Lancet* 265.6791, pp. 850–852.

Smeets, D., P. Bruyninckx, J. Keustermans, D. Vandermeulen, and P. Suetens (2010). "Robust Matching of 3D Lung Vessel Trees". In: *The Third International Workshop on Pulmonary Image Analysis*, pp. 61–70.

Sørensen, L., P. Lo, J. Petersen, A. Dirksen, and M. de Bruijne (2011). "Dissimilarity-based classification of anatomical tree structures". In: *Information Processing in Medical Imaging*. Ed. by G. Szkely and H. Hahn. Lecture Notes in Computer Science. Springer.

Stolk, J., R. A. Stockley, B. C. Stoel, B. G. Cooper, E. Piitulainen, N. Seersholm, K. R. Chapman, J. G. Burdon, M. Decramer, R. T. Abboud, G. P. Mannes, E. F. Wouters, J. E. Garrett, J. C. Barros-Tizon, E. W. Russi, D. A. Lomas, W. A. MacNee, and A. Rames (2012). "Randomised controlled trial for emphysema with a selective agonist of the gamma-type retinoic acid receptor". In: *Eur Respir J* 40.2, pp. 306–12.

Sun, S., C. Bauer, and R. Beichel (2012). "Automated 3-D Segmentation of Lungs With Lung Cancer in CT Data Using a Novel Robust Active Shape Model Approach". In: *IEEE Trans Med Imag* 31.2, pp. 449–460.

Tschirren, J., E. A. Hoffman, G. McLennan, and M. Sonka (2005a). "Intrathoracic airway trees: segmentation and airway morphology analysis from low-dose CT scans". In: *IEEE Trans Med Imag* 24.12, pp. 1529–39.

Tschirren, J., G. McLennan, K. Palágyi, E. A. Hoffman, and M. Sonka (2005b). "Matching and anatomical labeling of human airway tree". In: *IEEE Trans Med Imag* 24.12, pp. 1540–1547.

Washko, G. R., M. T. Dransfield, R. S. J. Estépar, A. Diaz, S. Matsuoka, T. Yamashiro, H. Hatabu, E. K. Silverman, W. C. Bailey, and J. J. Reilly (2009). "Airway wall attenuation: a biomarker of airway disease in subjects with COPD". In: *Journal of Applied Physiology* 107, pp. 185–191.

Weibel, E. R. (2009). "What makes a good lung?" In: *Swiss Med Wkly* 139.27-28, pp. 375–86.

Weinheimer, O., T. Achenbach, C. Bletz, C. Düber, H.-U. Kauczor, and C. P. Heussel (2008). "About Objective 3-D Analysis of Airway Geometry in Computerized Tomography". In: *IEEE Trans Med Imag* 27.1, pp. 64–74.

Wielputz, M. O., M. Eichinger, O. Weinheimer, S. Ley, M. A. Mall, M. Wiebel, A. Bischoff, H. U. Kauczor, C. P. Heußel, and M. Puderbach (2013). "Automatic airway analysis on multidetector computed tomography in cystic fibrosis: correlation with pulmonary function testing". In: *J Thorac Imaging* 28.2, pp. 104–13.

Wille, M. M., L. H. Thomsen, J. Petersen, S. B. Shaker, A. Dirksen, and J. H. Pedersen (2012). "Interobserver variability in visual evaluation of thoracic CT scans and comparison with automatic computer measurements of CT lung density". In: *European Respiratory Society, Annual Congres, Vienna*.

Wille, M. M., J. Petersen, A. Dirksen, and M. de Bruijne (2013). "Airway distensibility in chronic obstructive pulmonary disease - evaluation by CT airway segmentation and lung density measurement based on the Danish Lung Cancer Screening Trial". In: *American Thoracic Society International Conference (ATS)*.

Wu, X. and D. Chen (2002). "Optimal Net Surface Problems with Applications". In: *Automata, Languages and Programming*. Ed. by P. Widmayer, S. Eidenbenz, F. Triguero, R. Morales, R. Conejo, and M. Hennessy. Vol. 2380. LNCS. Springer, pp. 775–775.

Yin, Y., Q. Song, and M. Sonka (2009). "Electric Field Theory Motivated Graph Construction for Optimal Medical Image Segmentation". In: *Graph-Based Representations in Pattern Recognition*. Ed. by A. Torsello, F. Escolano, and L. Brun. Vol. 5534. LNCS. Springer, pp. 334–342.

Zheng, Y., K. Steiner, T. Bauer, J. Yu, D. Shen, and C. Kambhamettu (2007). "Lung Nodule Growth Analysis from 3D CT Data with a Coupled Segmentation and Registration Framework". In: Los Alamitos, CA, USA: IEEE Computer Society, pp. 1–8.