The Second International Workshop on Pulmonary Image Analysis

London, September 20, 2009

EDITED BY

Matthew Brown
Marleen de Bruijne
Bram van Ginneken
Atilla Kiraly
Jan-Martin Kuhnigk
Cristian Lorenz
Jamie McClelland
Kensaku Mori
Anthony Reeves
Joseph Reinhardt

ISBN-13: 978-1-4486-8089-1

Preface

After the successful first edition of the International Workshop on Pulmonary Image Analysis at MICCAI 2008 in New York City, the entire organizing team volunteered to organize the second edition of this event, aimed at bringing together researchers in pulmonary image analysis to discuss recent advances in this rapidly developing field. The Second International Workshop on Pulmonary Image Analysis will be held on September 20, 2009 in London, UK, again as a workshop of the MICCAI conference. Two researchers later joined the organizing team. We received many high quality submissions for this workshop. All papers underwent a thorough review process with two to four reviews per paper by members of the program committee and additional reviewers.

The proceedings of this workshop consist of three parts. There are fifteen regular papers, dealing with various aspects of image analysis of pulmonary image data, including segmentation, registration, and quantification of abnormalities in various modalities, with the focus in most studies on computed tomography, but also with papers on the analysis of MRI and X-ray scans. Next to these regular papers, we invited researchers to join in two comparative studies where algorithms were applied to a common data set, and submit a paper to the workshop about their system. The first of these challenges is EXACT09, on the extraction of the pulmonary airway tree from CT data. The second one, VOLCANO'09, is on the analysis of size changes in pulmonary nodules from consecutive CT scans. The results of these challenges are described in two overview papers that can be found in these proceedings. Moreover, fifteen papers describe systems that participated in the EXACT09 challenge and three papers describe algorithms that were used for the VOLCANO'09 challenge. That challenge attracted thirteen participating teams who applied algorithms, often previously published and not described in these proceedings, to the challenge data.

Challenges have been held at MICCAI since 2007 when the first edition of the Grand Challenge workshop compared approaches for liver segmentation in CT and caudate segmentation in MRI data. The challenge workshops have proved popular and allow a direct comparison between different approaches to solve the same problem. They have been organized in conjunction with other conferences as well, and one of these, ANODE09, was tailored particularly towards pulmonary image analysis. EXACT09 and VOLCANO'09 are the first challenges organized at MICCAI outside of the Grand Challenge workshops. The collection of good data set is paramount for such comparisons and for EXACT09 this was only possible through the help of most members of the organizing team who used their extensive networks to collect a diverse set of scans. The VOLCANO'09 challenge was only possible through the unique data made available by Weill Medical College of Cornell University.

We would like to take this opportunity to thank the MICCAI 2009 organizers for their excellent organizational support, and all the reviewers for helping us with the paper selection. We acknowledge the generous contributions of Siemens Corporate Research, Philips Medical Systems, VIDA Diagnostics, and MeVis Medical Solutions which helped make this event possible.

Matthew Brown
Marleen de Bruijne
Bram van Ginneken
Atilla Kiraly
Jan-Martin Kuhnigk
Cristian Lorenz
Jamie R. McClelland
Kensaku Mori
Anthony P. Reeves
Joseph M. Reinhardt

Organization

Organizers and Program Committee

Matthew Brown *(UCLA Radiological Sciences, Los Angeles, USA)*
Marleen de Bruijne *(Department of Computer Science, University of Copenhagen, Denmark; Biomedical Imaging Group Rotterdam, Erasmus MC, Rotterdam, The Netherlands)*
Bram van Ginneken *(Image Sciences Institute, UMC Utrecht, The Netherlands)*
Atilla Kiraly *(Siemens Corporate Research, Princeton, USA)*
Jan-Martin Kuhnigk *(Fraunhofer MEVIS, Bremen, Germany)*
Cristian Lorenz *(Philips Research Europe, Hamburg, Germany)*
Jamie McClelland *(Centre of Medical Image Computing, University College London, London, United Kingdom)*
Kensaku Mori *(Department of Media Science, Graduate School of Information Science, Nagoya University, Japan)*
Anthony Reeves *(School of Electrical and Computer Engineering, Cornell University, Ithaca, USA)*
Joseph Reinhardt *(Department of Biomedical Engineering, The University of Iowa, Iowa City, USA)*

Reviewers

Alberto Biancardi
Tobias Böhler
Matt Brown
Marleen de Bruijne
Kunlin Cao
Kai Ding
Sergei Fotin
Bram van Ginneken
Vladlena Gorbunova
Anja Hennemuth
Jing Huo
Artit Jirapatnakul
Atilla Kiraly
Tobias Klinder
Jan-Martin Kuhnigk
Jaesung Lee
Pechin Lo
Jamie McClelland
Keelin Murphy
Anthony Reeves
Joseph Reinhardt
Eva van Rikxoort
Michael Schmidt
Youbing Yin

Sponsoring

This workshop is sponsored in part by contributions from MeVis Medical Solutions, Philips Medical Systems, Siemens Corporate Research, and VIDA Diagnostics.

Table of Contents

Regular Workshop Papers

Oral Presentations

Poster Presentations

EXACT'09 Challenge Papers

Oral Presentations

Poster Presentations

VOLCANO '09 Challenge Papers

Oral Presentations

Poster Presentations

Regular Workshop Papers

Prediction of Respiratory Motion Using A Statistical 4D Mean Motion Model

Jan Ehrhardt[1], René Werner[1], Alexander Schmidt–Richberg[1], and Heinz Handels[1]

Department of Medical Informatics, University Medical Center Hamburg–Eppendorf, Germany, j.ehrhardt@uke.uni-hamburg.de

Abstract. In this paper we propose an approach to generate a 4D statistical model of respiratory lung motion based on thoracic 4D CT data of different patients. A symmetric diffeomorphic intensity–based registration technique is used to estimate subject–specific motion models and to establish inter–subject correspondence. The statistics on the diffeomorphic transformations are computed using the Log–Euclidean framework. We present methods to adapt the genererated statistical 4D motion model to an unseen patient–specific lung geometry and to predict individual organ motion. The prediction is evaluated with respect to landmark and tumor motion. Mean absolute differences between model–based predicted landmark motion and corresponding breathing–induced landmark displacements as observed in the CT data sets are 3.3 ± 1.8 *mm* considering motion between end expiration to end inspiration, if lung dynamics are not impaired by lung disorders.
The statistical respiratory motion model presented is capable of providing valuable prior knowledge in many fields of applications. We present two examples of possible applications in the fields of radiation therapy and image guided diagnosis.

1 Introduction

Respiration causes significant motion of thoracical and abdominal organs and thus is a source of inaccuracy in image guided interventions and in image acquisition itself. Therefore, modeling and prediction of breathing motion has become an increasingly important issue within many fields of application, e.g in radiation therapy [1].

Based on 4D images, motion estimation algorithms enable to determine patient–specific spatiotemporal information about movements and organ deformation during breathing. A variety of respiratory motion estimation approaches have been developed in the last years, ranging from using simple analytical functions to describe the motion over landmark–, surface– or intensity–based registration techniques [2, 3] to biophysical models of the lung [4]. However, the computed motion models are based on individual 4D image data and their use is usually confined to motion analysis and prediction of an individual patient.

The key contribution of this article is the generation of a *statistical* 4D inter–individual motion model of the lung. A symmetric diffeomorphic non–linear

intensity–based registration algorithm is used to estimate lung motion from a set of 4D CT images from different patients acquired during free breathing. The computed vector motion fields are transformed into a common coordinate system and a 4D mean motion model (4D–MMM) of the respiratory lung motion is extracted using the Log–Euclidean framework [5] to compute statistics on the diffeomorphic transformations. Furthermore, methods are presented to adapt the computed 4D–MMM to the patient's anatomy in order to predict individual organ motion without 4D image information. We perform a quantitative in–depth evaluation of the model–based prediction accuracy for intact and impaired lungs and two possible applications of the 4D–MMM in the fields of radiation therapy and image guided diagnosis are shown.

Few works that deal with the development of statistical lung motion models have been published. Some approaches exist for the generation of 3D lung atlases [6], or the geometry–based simulation of cardiac and respiratory motions [7]. First steps towards an average lung motion model generated from different patients were done by Sundaram et al. [8], but their work focuses on 2D+t lung MR images and the adaptation of the breathing model to a given patient has not been addressed. First methods for building inter–patient models of respiratory motion and the utilization of the generated motion model for model–based prediction of individual breathing motion were presented in [9] and [10]. This paper is an extension of [10] with regard to the methodology and the quantitative evaluation. In [9] motion models were generated by applying a Principal Component Analysis (PCA) to motion fields generated by a surface–based registration in a population of inhale–exhale pairs of CT images. Our approach is different in all aspects: the registration method, the solution of the correspondance problem, the spatial transformation of motion fields, and the computation of statistics of the motion fields. Furthermore, we present a detailed quantitative evaluation of a model based prediction for intact and impaired lungs. This offers interesting insights into the prediction accuracy to be expected depending on size and position of lung tumors.

2 Method

The goal of our approach is to generate a statistical model of the respiratory lung motion based on a set of N_p thoracic 4D CT image sequences. Each 4D image sequence is assumed to consist of N_j 3D image volumes $I_{p,j} : \Omega \rightarrow \mathbb{R}$ ($\Omega \subset \mathbb{R}^3$), which are acquired at corresponding states of the breathing cycle. This correspondance is ensured by the applied 4D image reconstruction method [11] and therefore, a temporal alignment of the patient data sets is not necessary.

Our method consists of three main steps: First, the subjectspecific motion is estimated for each 4D image sequence by registering the 3D image frames. In a second step, an average shape and intensity model is generated from the CT images. In the last step, the average shape and intensity model is used as anatomical reference frame to match all subject-specific motion models and to build an average intersubject model of the respiratory motion.

Image registration is required in all three steps. We use a non–linear, intensity–based, diffeomorphic registration method as described in the next section. The three steps to generate the statistical model of the respiratory motion are detailed in Sect. 2.2. The utilization of the 4D–MMM for motion predicition is presented in Sect. 2.3.

2.1 Diffeomorphic image registration

Diffeomorphic mappings $\boldsymbol{\varphi} : \Omega \rightarrow \Omega$, $(\boldsymbol{\varphi} \in Diff(\Omega),\ \Omega \subset \mathbb{R}^d)$ guarantee that the topology of the transformed objects is preserved and are therefore used in computational anatomy to analyze and characterize the biological variability of human anatomy [12]. A practical approach for fast diffeomorphic image registration was recently proposed in [13] by constraining $\boldsymbol{\varphi}$ to a subgroup of diffeomorphisms. Here, diffeomorphisms are parametrized by a stationary velocity field $\boldsymbol{v}$, and the diffeomorphic transformation $\boldsymbol{\varphi}$ is given by the solution of the *stationary* flow equation at time $t = 1$ [5]:

$$\frac{\partial}{\partial t}\phi(\boldsymbol{x}, t) = \boldsymbol{v}(\phi(\boldsymbol{x}, t)) \text{ and } \phi(\boldsymbol{x}, 0) = \boldsymbol{x}. \tag{1}$$

The solution of eq. (1) is given by the group exponential map $\boldsymbol{\varphi}(\boldsymbol{x}) = \phi(\boldsymbol{x}, 1) = \exp(\boldsymbol{v}(\boldsymbol{x}))$ and the significant advantage of this approach is that these exponentials can be computed very efficiently (see [5] for details).

The problem of image registration can now be understood as finding a parametrizing velocity field $\boldsymbol{v}$, so that the diffeomorphic transformation $\boldsymbol{\varphi} = \exp(\boldsymbol{v})$ minimizes a distance $\mathcal{D}$ between a reference image I_0 and the target image I_j with respect to a desired smoothness $\mathcal{S}$ of the transformation: $\mathcal{J}[\boldsymbol{\varphi}] = \mathcal{D}[I_0, I_j; \boldsymbol{\varphi}] + \alpha\mathcal{S}[\boldsymbol{\varphi}]$. Using $\mathcal{S}[\boldsymbol{\varphi}] = \int_\Omega \|\nabla \boldsymbol{v}\|^2 d\boldsymbol{x}$ (with $\boldsymbol{\varphi} = \exp(\boldsymbol{v})$) as regularization scheme, the following iterative registration algorithm can be derived:

Algorithm 1 Symmetric diffeomorphic registration

Set $\boldsymbol{v}^0 = 0$, $\boldsymbol{\varphi} = \boldsymbol{\varphi}^{-1} = Id$ and $k = 0$
repeat
  Compute the update step $\boldsymbol{u} = \frac{1}{2}\left(\boldsymbol{f}_{I_0, I_j \circ \boldsymbol{\varphi}} - \boldsymbol{f}_{I_j, I_0 \circ \boldsymbol{\varphi}^{-1}}\right)$
  Update the velocity field and perform a diffusive regularization:

$$\boldsymbol{v}^{k+1} = (Id - \tau\alpha\boldsymbol{\Delta})^{-1}\left(\boldsymbol{v}^k + \tau\boldsymbol{u}\right) \tag{2}$$

  Calculate $\boldsymbol{\varphi} = \exp(\boldsymbol{v}^{k+1})$ and $\boldsymbol{\varphi}^{-1} = \exp(-\boldsymbol{v}^{k+1})$
  Let $k \leftarrow k + 1$
until $\|\boldsymbol{v}^{k+1} - \boldsymbol{v}^k\| < \epsilon$ or $k \geq K_{max}$

The update field $\boldsymbol{u}$ is calculated in an inverse consistent form to assure source to target symmetry. The force term $\boldsymbol{f}$ is related to $\mathcal{D}$ and is chosen to be:

$$\boldsymbol{f}_{I_0, I_j \circ \boldsymbol{\varphi}}(\boldsymbol{x}) = -\frac{(I_0(\boldsymbol{x}) - (I_j \circ \boldsymbol{\varphi})(\boldsymbol{x}))\, \nabla (I_j \circ \boldsymbol{\varphi})(\boldsymbol{x})}{\|\nabla (I_j \circ \boldsymbol{\varphi})(\boldsymbol{x})\|^2 + \kappa^2 (I_0(\boldsymbol{x}) - (I_j \circ \boldsymbol{\varphi})(\boldsymbol{x}))^2} \tag{3}$$

with κ^2 being the reciprocal of the mean squared spacing. Eq. (2) performs the update of the velocity field $\boldsymbol{v}$, where τ is the step width. The term $(Id - \tau\alpha\boldsymbol{\Delta})^{-1}$ is related to the diffusive smoother $\mathcal{S}$ and can be computed efficiently using additive operator splitting (AOS).

We have chosen this diffeomorphic registration approach because of three reasons: In the context of the motion model generation, it is important to ensure that the calculated transformations are symmetric and diffeomorphic because of the multiple usage of inverse transformations. The second reason is related to runtime and memory requirements: due to the size of the 4D CT images diffeomorphic registration algorithms using non–stationary vector fields, e.g. [14], are not feasible. Third, the representation of diffeomorphic transformations by stationary vector fields provides a simple way for computing statistics on diffeomorphisms via vectorial statistics on the velocity fields.

For a diffeomorphism $\boldsymbol{\varphi} = \exp(\boldsymbol{v})$, we call the velocity field $\boldsymbol{v} = \log(\boldsymbol{\varphi})$ the logarithm of $\boldsymbol{\varphi}$. Remarkably, the logarithm $\boldsymbol{v} = \log(\boldsymbol{\varphi})$ is a simple 3D vector field and this allows to perform vectorial statistics on diffeomorphisms, while preserving the invertibility constraint [15]. Thus, the Log-Euclidean mean of diffemorphisms is given by averaging the parametrizing velocity fields:

$$\bar{\boldsymbol{\varphi}} = \exp\left(\frac{1}{N}\sum_i \log(\boldsymbol{\varphi}_i)\right). \tag{4}$$

The mean and the distance are inversion-invariant, since $\log(\boldsymbol{\varphi}) = -\log(\boldsymbol{\varphi}^{-1})$. Even though the metric linked to this distance is not translation invariant, it provides a powerful framework where statistics can be computed more efficiently than in the Riemannian distance framework. For a more detailed introduction to the mathematics of the diffeomorphism group and the associated tangent space algebra, we refer to [5] and the references therein.

2.2 Generation of a 4D mean motion model

In the first step, we estimate the intra–patient respiratory motion for each 4D image sequence by registering the 3D image frames. Let $I_{p,j} : \Omega \to \mathbb{R}$ ($\Omega \subset \mathbb{R}^3$) be the 3D volume of subject $p \in \{1, \ldots, N_p\}$ acquired at respiratory state $j \in \{0, \ldots, N_j - 1\}$. Maximum inhale is chosen as reference breathing state and the diffeomorphic transformations $\boldsymbol{\varphi}_{p,j} : \Omega \to \Omega$ are computed by registering the reference image $I_{p,0}$ with the target images $I_{p,j}$, $j \in \{1, \ldots, N_j - 1\}$. In order to handle discontinuities in the respiratory motion between pleura and rib cage, lung segmentation masks are used to restrict the registration to the lung region by computing the update field only inside the lung (see [3] for details).

In order to build a statistical model of respiratory motion, correspondence between different subjects has to be established, i.e. an anatomical reference frame is necessary. Therefore, the reference images $I_{p,0}$ for $p = 1, \ldots, N_p$ are used to generate an average intensity and shape atlas $\bar{I}_0$ of the lung in the reference breathing state by the method described in [10]. This 3D atlas image $\bar{I}_0$ is now used as reference frame for the statistical lung motion model. Each patient–specific reference image $I_{p,0}$ is mapped to the average intensity and shape atlas $\bar{I}_0$ by an affine alignment and a subsequent diffeomorphic registration.

Let $\boldsymbol{\psi}_p$ be the transformation between the reference image $I_{p,0}$ of subject p and the atlas image $\bar{I}_0$. Since the intra–subject motion models $\boldsymbol{\varphi}_{p,j}$ are defined in the anatomical spaces of $I_{p,0}$, we apply a coordinate transformation

$$\tilde{\boldsymbol{\varphi}}_{p,j} = \boldsymbol{\psi}_p \circ \boldsymbol{\varphi}_{p,j} \circ \boldsymbol{\psi}_p^{-1} \tag{5}$$

to transfer the intra–subject deformations into the atlas coordinate space. Such a coordinate transformation accounts for the differences in the coordinate systems of subject and atlas due to misalignment and size/shape variation and eliminates subject–specific size, shape and orientation information in the deformation vectors. This enables the motion fields of each of the subjects to be compared directly quantitatively and qualitatively and the 4D–MMM is generated by calculating the Log-Euclidean mean $\bar{\boldsymbol{\varphi}}_j$ of the mapped transformations for each breathing state j:

$$\bar{\boldsymbol{\varphi}}_j = \exp\left(\frac{1}{N_p}\sum_p \log\left(\tilde{\boldsymbol{\varphi}}_{p,j}\right)\right) = \exp\left(\frac{1}{N_p}\sum_p \log\left(\boldsymbol{\psi}_p \circ \boldsymbol{\varphi}_{p,j} \circ \boldsymbol{\psi}_p^{-1}\right)\right). \tag{6}$$

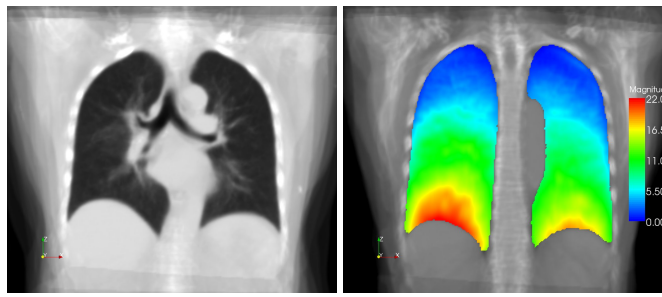
The method proposed in [16] was used to compute the logarithms $\log\left(\tilde{\boldsymbol{\varphi}}_{p,j}\right)$.

The resulting 4D–MMM consists of an average lung image $\bar{I}_0$ for a reference state of the breathing cycle, e.g. maximum inhalation, and a set of motion fields $\bar{\boldsymbol{\varphi}}_j$ describing an average motion between the respiratory state j and the reference state (Fig. 1).

2.3 Utilization of the 4D–MMM for individual motion prediction

The 4D–MMM generated in section 2.2 can be used to predict respiratory lung motion of a subject s even if no 4D image information is available. Presuming a 3D image $I_{s,0}$ acquired at the selected reference state of the breathing cycle is available, the 4D–MMM is adapted to the individual lung geometry of subject s by registering the average lung atlas $\bar{I}_0$ with the 3D image $I_{s,0}$. The resulting transformation $\boldsymbol{\psi}_s$ is used to apply the coordinate transformation eq. 5 to the mean motion fields $\bar{\boldsymbol{\varphi}}_j$ in order to obtain the model–based prediction of the subject–specific lung motion: $\hat{\boldsymbol{\varphi}}_{s,j} = \boldsymbol{\psi}_s^{-1} \circ \bar{\boldsymbol{\varphi}}_j \circ \boldsymbol{\psi}_s$.

However, two problems arise. First, breathing motion of different individuals varies significantly in amplitude [1]. Therefore, motion prediction using the mean amplitude will produce unsatisfying results. To account for subject–specific motion amplitudes, we propose to introduce additional information by providing the



(a) (b)

Fig. 1. Visualization of average lung model $\bar{I}_0$ (a) and magnitude of mean deformation $\bar{\varphi}_j$ between end inspiration and end expiration (b). The average deformation model shows a typical respiratory motion pattern. Different windowing and leveling functions are used to accentuate inner/outer lung structures.

required change in lung air content ΔV_{air}. Even without 4D-CT data, this information can be acquired by spirometry measurements. Thus, we search a scaling factor λ so that the air content of the transformed reference image $I_{s,0} \circ \lambda \hat{\boldsymbol{\varphi}}_{s,j}^{-1}$ is close to the air content $V_{air}\left(I_{s,0}\right) + \Delta V_{air}$. In order to ensure that the scaled motion field is diffeomorphic, the scaling is performed in the Log–Euclidean. To determine the correct scaling factor λ, a binary search strategy is applied and the air content is computed using the method described in [17]. ΔV_{air} can be regarded as a parameter that describes the depth of respiration. In general, other measurements can also be used to calculate appropriate scaling factors, e.g. the amplitude of the diaphragm motion.

Further, a second problem arises when predicting individual breathing motion of lung cancer patients. Lung tumors will impair the atlas–patient registration because there is no corresponding structure in the atlas. This leads to distortions in $\boldsymbol{\psi}_s$ near the tumor region and consequently the predicted motion fields $\hat{\boldsymbol{\varphi}}_{s,j}$ are affected. Therefore, we decided to compute $\boldsymbol{\psi}_s$ by registering lung segmentation masks from atlas and subject s and by omitting the inner lung structures.

3 Results

To capture the respiratory motion of the lung, 18 4D CT images were acquired using a 16–slice CT scanner operating in cine-mode. The scanning protocol and optical–flow based reconstruction method was described in [11]. The spatial resolution of the reconstructed 4D CT data sets is between $0.78 \times 0.78 \times 1.5mm^3$ and $0.98 \times 0.98 \times 1.5mm^3$. Each data set consists of 3D CT images at 10 to 14 preselected breathing phases. Due to computation times, in this study we use the following 4 phases of the breathing cycle: end inspiration (EI), 42% exhale (ME), end expiration (EE) and 42% inhale (MI). A clinical expert delineated left and right lung and the lung tumors in the images.

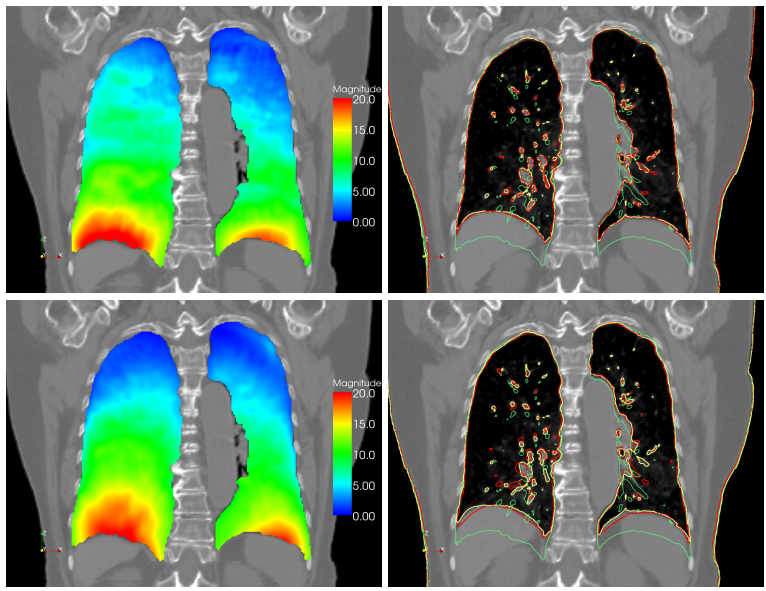


Fig. 2. Result of the motion estimatation by intra–patient registration (top row) and the model–based motion prediction (bottom row) of patient 01. Visualization of the magnitude of the displacement field computed by intra–patient registration (top left) and of the displacement field predicted by the 4D mean motion model (bottom left). Right: contours at end inspiration (green), end expiration (yellow) and estimated/predicted contours at end expiration (red).

The aim of the model generation is to create a representation of the mean healthy lung motion. In a dynamic MRI study by Plathow et al. [18], tumors with diameter $> 3cm$ were shown to influence respiratory lung dynamics. According to their observations, we divide the lungs into two groups: lungs with *intact* dynamics and lungs with *impaired* motion. Lungs without or with only small tumors (volume $< 14.1cm^3$ or diameter $< 3cm$) are defined as intact. Lungs with large tumors or lungs affected by other diseases (e.g. emphysema) are defined as impaired. According to this partitioning, we have 12 data sets with both lungs intact and 6 data sets with at least one impaired lung. Only data sets with intact lungs are used to generate the 4D–MMM.

3.1 Landmark–based evaluation

Due to the high effort of the manual landmark identification only 10 of the 18 data sets are used for the detailed quantitative landmark–based evaluation. Between 70 and 90 inner lung landmarks (prominent bifurcations of the bronchial tree and the vessel tree) were identified manually in the four breathing phases, about 3200 landmarks in total. An intraobserver variability of $0.9 \pm 0.8mm$ was

Table 1. Landmark motion amplitudes and target registration errors $\overline{R}_{EE}$ for the patients considered (in *mm*). Values are averaged over all landmarks per lung. Lungs with impaired motion are indicated by a gray text color.

Data set	(Lung)	Landmark motion [mm]	Intra-patient registration TRE [mm]	Model-based prediction TRE [mm]
Patient01	left	4,99 ± 4,84	1,51 ± 1,31	2.43 ± 1,64
	right	7,25 ± 4,47	1,41 ± 0,83	3.97 ± 2,08
Patient02	left	7,09 ± 2,92	2,28 ± 1,73	4.26 ± 1,28
	right	4,21 ± 1,75	1,16 ± 0,61	3.82 ± 1,14
Patient03	left	6,15 ± 2,26	1,38 ± 0,73	3.68 ± 1,31
	right	6,28 ± 2,01	1,78 ± 1,05	3.72 ± 1,37
Patient04	left	6,65 ± 2,56	1,53 ± 0,93	4.01 ± 1,60
	right	6,22 ± 3,52	1,44 ± 0,82	2.28 ± 1,09
Patient05	left	5,77 ± 2,03	1,50 ± 0,80	3.17 ± 1,34
	right	3,18 ± 3,36	1,29 ± 1,04	3.47 ± 1,99
Patient06	left	9,67 ± 8,32	1,64 ± 1,42	5.85 ± 2,65
	right	11,85 ± 7,08	1,60 ± 1,00	4.88 ± 2,02
Patient07	left	8,22 ± 6,52	2,45 ± 2,22	3.99 ± 1,79
	right	4,99 ± 6,65	1,49 ± 1,48	3.35 ± 1,69
Patient08	left	5,78 ± 4,14	1,18 ± 0,57	3.15 ± 1,70
	right	6,28 ± 5,63	1,25 ± 1,03	3.11 ± 2,24
Patient09	left	7,43 ± 5,34	1,42 ± 1,22	3.05 ± 1,39
	right	8,41 ± 5,22	1,67 ± 1,03	4.94 ± 3,01
Patient10	left	7,63 ± 5,83	1,93 ± 2,10	3.16 ± 2,29
	right	8,85 ± 6,76	1,76 ± 1,33	5.12 ± 2,34

assessed by repeated landmark identification in all test data sets. The target registration error (TRE) was determined for a quantitative evaluation of the patient–specific registration method and the model–based prediction. The TRE R_j^k is the difference between the motion of landmark k estimated by φ_j and the landmark motion as observed by the medical expert.

The mean landmark motion magnitude, i.e. the mean distance of corresponding landmarks, between EI and EE is $6.8 \pm 5.4mm$, ($2.6 \pm 1.6mm$ between EI and ME and $5.0 \pm 2.8mm$ between EI and MI). The TRE of the intra–patient registration is a lower bound for the accuracy of the model–based prediction using the 4D–MMM. The average TRE $\overline{R}_{EE}$ between the reference phase (EI) and EE for patient 01 to 10 (averaged over all landmarks and patients) is $1.6 \pm 1.3mm$ ($1.5 \pm 0.8mm$ between EI and ME and $1.6 \pm 0.9mm$ between EI and MI). Details for all test data sets are shown in table 1.

For each of the 10 test data sets the 4D–MMM is used to predict landmark motion as described in Sect. 2.3. If both lungs of the test data set are intact, a leave–one–out strategy is applied to ensure that the patient data is not used for the model generation. The change in lung air content ΔV_{air} needed for the computation of the scaling factor λ was calculated from the CT images I_{EI} and I_{EE} for each lung side and each test data set. The same factor λ was used to

scale the predicted motion fields $\hat{\varphi}_{EE}$, $\hat{\varphi}_{ME}$ and $\hat{\varphi}_{MI}$. Besides ΔV_{air} no 4D information is used for the model–based prediction.

In Fig. 2 the motion field predicted by the 4D–MMM is compared to the motion field computed by patient–specific registration. A good correspondency between the motion fields is visible, except in the right upper lobe where small deviations occur. The prediction accuracy is illustrated by overlayed contours.

The average TREs $\overline{R}_{EE}$ are listed in table 1 for each of the test data sets and for both the patient–specific and model–based motion estimation. Lungs with impaired motion are indicated by a gray text color. Regarding table 1, lungs with impaired motion generally show higher TREs for the model–based prediction than intact lungs. The average TRE $\overline{R}_{EE}$ for intact lungs is $3.3 \pm 1.8mm$, which is significantly lower ($p < 0.01$) than for lungs with impaired motion ($\overline{R}_{EE} = 4.2 \pm 2.2mm$). Significance is tested by applying a multilevel hierarchical model with the individual R^k values nested within the patient (software: SPSS v.17); data are logarithmized to ensure normal distribution and the model is adjusted to landmark motion.

3.2 Model-based prediction of tumor motion

For a second evaluation of the model, we use expert generated tumor segmentations in two breathing phases (EI and EE) of 9 patient data sets with solid lung tumors. The 4D–MMM is transformed into the coordinate space of each test data set (see Sect. 2.3) and then used to warp the expert–generated tumor segmentation at maximum exhale towards maximum inhale. The distance between the predicted tumor mass center and the center of the manual segmentation was used to evaluate the accuracy of the model–based prediction. Corresponding results are summarized in table 2. Large tumors with a diameter $> 3cm$ are marked in the table as "large".

Regarding table 2 accuracy of the model–based predicted motion of the tumor mass center from EI to EE ranges from $0.66mm$ to $7.38mm$. There is no significant correlation between the tumor motion amplitude and the accuracy of the model-based predicted mass center ($r = 0.19$, $p > 0.15$). Furthermore, it cannot be shown that the prediction accuracy for small tumors is significantly better than for large tumors ($p > 0.4$). In contrast, the model–based prediction accuracy of non-adherent tumors is significantly better than for tumors adhering to chest wall or hilum ($p < 0.05$). In these cases the model presumes the tumour moves like surrounding lung tissue, whereas it rather moves like the adjacent non-lung structure (e.g. chest wall or hilum). In the last column in table 2 those tumors are tagged. Significance is tested by applying a linear mixed model (software: SPSS v.17) and the model is adjusted to tumor motion.

4 Discussion

In this paper, we proposed a method to generate an inter–subject statistical model of the breathing motion of the lung, based on individual motion fields

Table 2. Tumor size and motion amplitude, and the center distances between manually segmented tumor and predicted tumor position (see text for details).

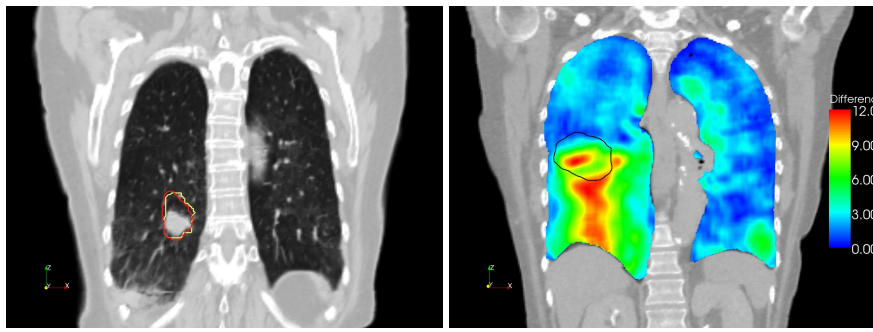
Data set	(Lung)	Tumor size [cm^3]	Tumor motion [mm]	Intra–patient registration TRE [mm]	Model–based prediction TRE [mm]	large	adhere
Patient 01	right	6.5	12.20	0.45	3.54		
Patient 02	right	7.6	2.15	1.44	3.90		X
Patient 03	left	12.7	6.74	0.41	3.91		X
Patient 05	right	8.2	2.34	1.95	5.39		X
	right	17.3	1.68	1.05	4.44	X	X
Patient 06	left	3.4	19.78	2.12	6.87		
	right	128.2	13.78	0.97	2.99	X	
Patient 07	right	2.8	1.31	0.42	0.66		
Patient 08	right	18.4	6.24	0.90	1.59	X	
Patient 09	right	88.9	8.35	0.29	5.33	X	X
Patient 10	right	96.1	1.77	1.01	7.46	X	X

extracted from 4D CT images. Methods to apply this model in order to predict patient–specific breathing motion without knowledge of 4D image information were presented. Ten 4D CT data sets were used to evaluate the accuracy of the image–based motion field estimation and the model–based motion field prediction. The intra–patient registration shows an average TRE in the order of the voxel size, e.g. $1.6 \pm 1.3mm$ when considering motion between EI and EE. The 4D–MMM achieved an average prediction error (TRE) for the motion between EI and EE of $3.3 \pm 1.8mm$. Regarding that besides the calculated scaling factor λ no patient–specific motion information is used for the model–based prediction and that the intra–patient registration as well as the atlas–patient registration is error prone, we think this is a promising result. Thus we believe that a statistical respiratory motion model has the capability of providing valuable prior knowledge in many fields of applications.

Since the statistical model represents intact respiratory dynamics, it was shown that the prediction precision is significantly lower for lungs affected by large tumours or lung disorders ($4.2 \pm 2.2mm$). These results indicate (at least for the 10 lung tumor patients considered) that large tumors considerably influence respiratory lung dynamics. This finding is in agreement with Plathow et al. [18]. In addition, we applied the 4D–MMM to predict patient–specific tumor motion. No correlation between prediction accuracy and tumor size or tumor motion amplitude could be detected (at least for our test data sets). We observed that tumors adhering to non–lung structures degrade local lung dynamics significantly and model–based prediction accuracy is decreased for these cases.

To conclude this paper, we present two examples of possible applications of the statistical respiratory motion model.

Application examples: The capability of the 4D–MMM to predict tumor motion for radiotherapy planning is exemplarily illustrated for patient 01. This patient has a small tumor not adherent to another structure, and a therapeutically



(a) (b)

Fig. 3. (a) Visualization of the internal target volume (ITV) of patient 01 in a coronal CT slice. The ITV was calculated from expert-defined tumor segmentations (yellow contour) and from tumor positions predicted by the average motion model (red contour). (b) Visualization of the difference between lung motion estimated by patient–specific registration and lung motion predicted by the 4D–MMM for patient 09. The left lung shows intact lung motion; dynamics of the right lung are impaired by the large tumor. The contour of the tumor is shown in black.

relevant tumor motion of 12.2mm. An important measure for planning in 3D conformal radiotherapy is the internal target volume (ITV), which contains the complete range of motion of the tumour. For this patient, the ITV is calculated first from expert-defined tumor segmentations in the images acquired at EI, EE, ME and MI. In a second step, the expert segmentation in EI is warped to EE, ME and MI using the 4D–MMM and the ITV is calculated based on the warped results. The outlines of both ITVs are shown in Fig. 3(a).

A second example demonstrates that the 4D–MMM could be helpful from the perspective of image-guided diagnosis. Here, the motion pattern of individual patients are compared to a "normal" motion, represented by the 4D–MMM. To visualize the influence of a large tumor to the respiratory motion, the difference between the individual motion field computed by intra–patient registration and the motion field predicted by the 4D–MMM is shown in Fig. 3(b). The left lung shows differences of only about 3mm, whereas the large differences to the intact lung motion indicate that the respiratory dynamics of the right lung are influenced by the large tumor.

Currently, the statistical motion model represents the average motion in the training population. A main focus of our future work is to include the variability of the motion into the model. Here, the Log–Euclidean framework provides a suitable technique for more detailed inter–patient statistics.

References

1. Keall, P.J., Mageras, G., Balter, J.M., et al.: The management of respiratory motion in radiation oncology report of AAPM task group 76. Med Phys **33**(10)

(2006) 3874–3900
2. Vik, T., Kabus, S., von Berg, J., Ens, K., Dries, S., Klinder, T., Lorenz, C.: Validation and comparison of registration methods for free-breathing 4D lung CT. In: SPIE Medical Imaging 2008. Volume 6914., SPIE (2008) 2P
3. Werner, R., Ehrhardt, J., Schmidt-Richberg, A., Handels, H.: Validation and comparison of a biophysical modeling approach and non-linear registration for estimation of lung motion fields in thoracic 4D CT data. In: SPIE Medical Imaging 2009: Image Processing. Volume 7259. (2009) 0U–1–8
4. Werner, R., Ehrhardt, J., Schmidt, R., Handels, H.: Patient-specific finite element modeling of respiratory lung motion using 4D CT image data. Med Phys **35**(5) (May 2009) 1500–1511
5. Arsigny, V.: Processing Data in Lie Groups: An Algebraic Approach. Application to Non-Linear Registration and Diffusion Tensor MRI. Thèse de sciences (phd thesis), École polytechnique (November 2006)
6. Li, B., Christensen, G.E., Hoffman, E.A., McLennan, G., Reinhardt, J.M.: Establishing a normative atlas of the human lung: intersubject warping and registration of volumetric CT images. Acad Radiol **10** (2003) 255–265
7. Segars, W., Lalush, D., Tsui, B.: Modeling respiratory mechanics in the MCAT and spline-based MCAT phantoms. IEEE Trans Nucl Sci **48**(1) (Feb 2001) 89–97
8. Sundaram, T.A., Avants, B.B., Gee, J.C.: A dynamic model of average lung deformation using capacity-based reparameterization and shape averaging of lung mr images. In: MICCAI 2004, Springer (2004) 1000–1007
9. Klinder, T., Lorenz, C., Ostermann, J.: Respiratory motion modeling and estimation. In: Workshop on Pulmonary Image Analysis, New York, USA (2008) 53–62
10. Ehrhardt, J., Werner, R., Schmidt-Richberg, A., Schulz, B., Handels, H.: Generation of a mean motion model of the lung using 4D CT data. In: Visual Computing for Biomedicine, Delft, Eurographics Association (2008) 69–76
11. Ehrhardt, J., Werner, R., Säring, D., Frenzel, T., Lu, W., Low, D., Handels, H.: An optical flow based method for improved reconstruction of 4D CT data sets acquired during free breathing. Med Phys **34**(2) (Feb 2007) 711–721
12. Miller, M.I.: Computational anatomy: shape, growth, and atrophy comparison via diffeomorphisms. NeuroImage **23 Suppl 1** (2004) S19–S33
13. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Symmetric log-domain diffeomorphic registration: a demons-based approach. In: Med Image Comput Comput Assist Interv, MICCAI 2008, Springer (2008) 754–761
14. Beg, M.F., Miller, M.I., Trouve, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int J Comp Vis **61**(2) (2005) 139–157
15. Arsigny, V., Commowick, O., Pennec, X., Ayache, N.: A log-Euclidean framework for statistics on diffeomorphisms. In: MICCAI 2006, Springer (2006) 924–931
16. Bossa, M.N., Gasso, S.O.: A new algorithm for the computation of the group logarithm of diffeomorphisms. In: MFCA 2008, New York, USA (2008)
17. Lu, W., Parikh, P.J., El Naqa, I.M., et al.: Quantitation of the reconstruction quality of a four-dimensional computed tomography process for lung cancer patients. Med Phys **32** (2005) 890–901
18. Plathow, C., Fink, C., Ley, S., M.Puderbach, M.Eichinger, I.Zuna, A.Schmähl, H.Kauczor: Measurement of tumor diameter-dependent mobility of lung tumors by dynamic MRI. Radiother Oncol **73** (2004) 349–354

Curve- and Surface-based Registration of Lung CT images via Currents

Vladlena Gorbunova[1], Stanley Durrleman[2,3], Pechin Lo[1], Xavier Pennec[2], and Marleen de Bruijne[1,4]

[1] Department of Computer Science, University of Copenhagen, Denmark
[2] Asclepios, INRIA Sophia Antipolis, France
[3] Centre de Mathematique et Leurs Application, ENS Cachan, France
[4] Biomedical Imaging Group Rotterdam, Erasmus MC, Rotterdam, the Netherlands

Abstract. Feature-based registration methods offer a robust alternative to intensity-based methods when intensities change because of pathology, image artifacts or differences in acquisition. For registration of lung CT images, we propose to use distinctive anatomical structures, such as the pulmonary vessel tree centerlines and lung surfaces, to establish correspondences between pairs of scans. In this respect, we develop and evaluate a curve- and surface-based registration method using currents. This method does not require point correspondence between structures. We conducted experiments on five pairs of images, where each pair consists of image volumes extracted at the end inhale and end exhale phases of a 4D-CT scan. To evaluate the registration, we used a set of 300 anatomical landmarks marked on each image pair. Using both vessel centerlines curves and lung surfaces yields better alignment (median error of 1.85 mm) than using only curves (2.37 mm) or surfaces (3.53 mm). The combined method achieves overall registration accuracy comparable to that of intensity-based registration, whereas the errors are made in different locations. This suggests that low dimensional geometrical features capture sufficient information to drive a reliable registration, while results can still be improved by combining intensity and feature based registration approaches into one framework.

1 Introduction

Registration of chest CT scans is an important subject within pulmonary image analysis. The general task of registration is to establish a point-to-point correspondence between two images. Registration of lung CT images can be used in various clinical applications, such as lung cancer radiotherapy planning and quantitative analysis of disease progression.

Image registration methods can be separated into two general groups: intensity-based and feature-based methods. Intensity-based methods integrate spatial information over the entire image domain, whereas feature-based methods require a representation of the image data in terms of distinctive geometrical structures. Feature-based methods offer more robust registration when image intensity is

changed, owning to for instance pathology, image artifacts or differences in scan protocol. Generally, segmentation of geometrical structures in lungs is less sensitive to intensity changes, since the method incorporates geometrical regularity constraints or prior anatomical knowledge. Moreover, segmentation of distinctive lung structures may be either corrected manually or delineated by a professional.

The most distinctive anatomical structures in lung CT images are vessels, airways, lobe fissures and lung surfaces. Lungs surface and lobe fissures define large-scale deformations of the lungs and provide an insight into the global motion of the lungs, while small-scale deformations are influenced by vessels and airway tree motion.

Feature-based registration relies on various geometrical structures, e.g., points, curves or surfaces. Thin-plate spline image registration [1–3] is the standard method for matching points under the assumption that deformations are small. For large deformations, a diffeomorphic point matching approach was developed by Joshi and Miller [4] and was later adapted for surface matching [5]. Current-based diffeomorphic method for surface matching under the large deformations, pioneered by Glaunès et. al. [5], was further developed and adapted for curve matching problem [6, 7]. Within a framework of currents, no point correspondence between structures is required.

Several surface-based registration methods were previously developed for lung CT images [8–10]. The outer surface of the lungs together with the outer surface of vessels were used in an algorithm similar to iterative closest point methods in [8]. Lung surface was used to register CT lung images [10] and to constrain intensity-based registration with a deformation field obtained from surface matching procedure [9]. The two main advantages of the feature-based registration of lung CT images via currents are: no point correspondence is required and unified representation of curves and surfaces. The low dimensional geometrical features, such as curves and surfaces contain much fewer points compared to dense intensity images, thus feature-based registration can be more efficient. Moreover, in the framework of currents, dimensionality of image features may be reduced even more without decreasing registration accuracy [11].

In this paper we apply the current-based registration method, pioneered by Glaunès et. al. [5] and further propagated by Durrleman [7, 12], to three feature sets: vessel centerlines, lung surface and combined set of centerlines and surface. We evaluated the registration methods on a set of 5 pairs of end exhalation and end inhalation phases of 4D-CT images with ground truth landmarks.

2 Registration via Currents

2.1 Representation of curves and surfaces

In the framework of currents [5, 6, 12], geometrical shapes such as curves and surfaces are represented with a set of vectors. A current is encoded with a finite set of vectors attached to the specified positions. A curve $C(x)$ can be defined with its tangent vector $\tau(x)$ at each position x. In a discrete setting, curve is

considered as a set of piece-wise linear segments where each segment is represented by its center point, tangential direction, and segment length. Similarly, a surface $S(x)$, with a given mesh, is represented with the normal direction $n(x)$, face center x and area. Both surfaces and curves are thus encoded into currents as a set of vectors. Geometrical shape in the framework of currents is defined in a weak form, as the action of the shape on a test vector field w from a space of possible vector fields W. The current of a curve $C(\omega)$ is defined by the path integral along the curve through the test vector field w,

$$C(\omega) = \int_C w(x)\tau(x)dx. \tag{1}$$

And the current of a surface $S(\omega)$ is defined by the flux of the vector field w trough the surface,

$$S(\omega) = \int_S w(x)n(x)dx. \tag{2}$$

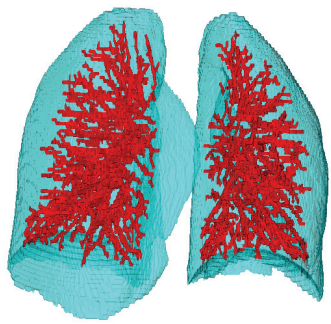
The space W of test vector fields is a space of square integrable vector fields convolved with a Gaussian kernel with standard deviation σ_w [12, 6]. The norm of the current, $\mu(C)$, is defined in the dual space W^*, as the maximum action of the current among all possible test vector fields $||\mu(C)||_W = \sup_{||\omega||_W \leq 1} C(w)$.

2.2 Lung structures as currents

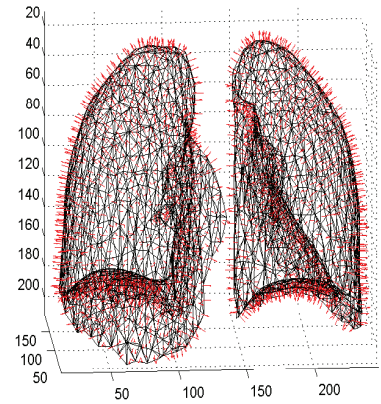
In this paper we used distinctive anatomical lung structures such as vessels and lung surface as features for registration. Fig. 1(a) shows an example of segmented lung structures. The lung fields and vessels are segmented with the algorithm described in [14]. A sparse triangulation of the lung surface was computed via the marching cube algorithm [15]. For each face, the corresponding normals were computed and oriented to point outwards of the surface. Fig. 1(b) shows an example of the constructed current for a lung surface.

Vessel tree was segmented as follows: lung image was thresholded with a fixed intensity value $t_v = -600HU$, then a local analysis of Hessian matrix was performed in order to remove non-tube like structures. Large vessels segmented near the hilum area were omitted from the vessel tree segmentation. For more details on vessels segmentation algorithm we refer the reader to [14]. Centerlines were extracted from the segmented vessel tree using a 3D thinning algorithm [16].

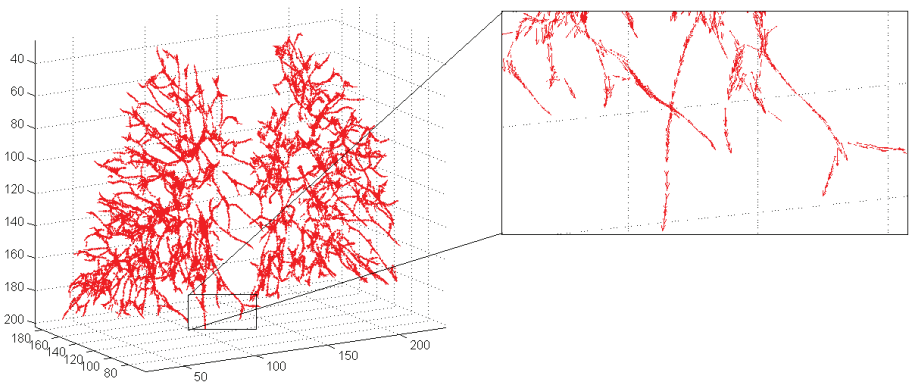
The tangential direction of a centerline was computed via local principal component analysis. For each centerline point we extracted neighboring centerline points, applied PCA to the point cloud, and assigned the first principal component to the tangential direction at the centerline. For centerlines sufficiently far from vessel bifurcation and neighboring vessel, the principal direction points to a tangential direction of the centerline. For centerlines close to the bifurcation the principal direction points between the two splitting vessel centerlines. This is consistent with the framework of currents, were the action of each vessel direction results in a joint action at the bifurcation point. The orientation for the



(a) Example of segmented lung surface and lung vessel tree



(b) Current corresponding to a lung surface.



(c) Current corresponding to a vessel tree centerlines.

Fig. 1. Example of segmented lungs surface and vessel tree 1(a); triangulation of the lungs surface (black mesh) with the corresponding current (red vectors) 1(b); current corresponding to the vessel tree centerlines (red vectors) with a zoom-in 1(c).

positive direction was set to point outwards from the center of the image. Fig. 1(c) shows an example of the constructed current for a segmented vessel tree and a zoom-in into a bottom part of the image.

2.3 Current-based Image Registration

In this paper, we combine the previous work on matching curves [6] and surfaces [5] via currents. The similarity measure between two curves C_f, C_m or

two surfaces S_f, S_m is defined as the squared norm of the difference in μ for corresponding currents with respect to the test vector field $w \in W$:

$$E(C_f; C_m) = ||\mu(C_f) - \mu(\phi(C_m))||^2_W, \tag{3}$$

for fixed and moving curves C_f and C_m respectively. And

$$E(S_f; S_m) = ||\mu(S_f) - \mu(\phi(S_m))||^2_W, \tag{4}$$

for fixed and moving surfaces S_f and S_m respectively, where ϕ is a diffeomorphic transform function. Combining two similarity terms for curves (3), surfaces (4) and a regularisation term with trade-off coefficients $\gamma_C, \gamma_S, \gamma_\phi$ in a final cost function gives:

$$\begin{aligned} E(C_f, S_f; C_m, S_m) = \gamma_C ||\mu(C_f) - \mu(\phi(C_m))||^2_W + \gamma_S ||\mu(S_f) - \mu(\phi(S_m))||^2_W \\ + \gamma_\phi Reg(\phi). \end{aligned} \tag{5}$$

Diffeomorphic transformation ϕ of curves and surfaces was modeled in the framework of large deformation diffeomorphic matching [4, 6], where deformation of each feature is defined by a velocity vector field $v_t = \phi'_t$. The smooth velocity field v_t is described via Gaussian kernel with standard deviation σ_V, where σ_V determines the typical scale of the deformations [12, 6]. To guarantee smoothness of the final diffeomorphism, we defined the regularisation term as in [12],

$$Reg(\phi) = \int_0^1 ||v_t||^2_V dt. \tag{6}$$

3 Experiments

In order to quantify the accuracy of the proposed registration method with a ground truth, we used images from a publicly available dataset [13]. For each image pair, 300 manually placed corresponding landmarks were provided [13]. Five pairs of images, where each pair consists of images extracted at end exhale and end inhale phases of 4D CT image, were used in our experiments. In-plane resolution of the images varied from 0.97×0.97 mm to 1.16×1.16 mm and slice thickness was 2.5 mm.

3.1 Parameter Settings

Vessel tree were segmented using the algorithm as in [14] with the intensity threshold -600 HU, ratio of Hessian eigenvalues was set to $m_1 = 0.75$, $m_2 = 0.5$. For every centerline point we extracted a neighboring centerline points from the cube neighborhood of $7 \times 7 \times 7$ voxels size and computed the principal direction of the centerlines. All the direction vectors were normalized to 1. Fig. 1(c) shows an example of the extracted currents for vessel centerlines with a zoom-in to a lower part of the lungs. A regular surface triangulation was constructed with a

marching cube algorithm with further simplification of the mesh [15]. Normal directions to each of the face were normalized to 1.

In our experiments, end inhale phase of 4D-CT image was registered to end exhale phase. The following internal parameters of image registration were selected manually. The accuracy of feature alignment σ_W was set to 5 mm for curves and 10 mm for surface features. The spatial variability of deformation velocity field σ_V was set to 25 mm for both types of features. The weight coefficients in the cost function (5) were set to $\gamma_C = 1$ for the curve matching term, $\gamma_S = 0.01$ for the surface matching term and $\gamma_\phi = 10^{-4}$ for the regularizer. The cost function was minimized with a standard gradient descent approach.

3.2 Results

We evaluated four registration methods, as follows: combined curve- and surface-based registration with cost function (5); curve-based registration with cost function (3); surface-based registration with cost function (4); and a free-form B-Spline intensity-based method as in [17]. We compared registration accuracy of the four methods based on the alignment of 300 landmarks distributed uniformly in lung area, Fig. 2(b) shows an example of the spatial distribution of landmarks within the lungs.

The overall accuracy of the image registration methods was defined as the mean Euclidean distance between landmarks, target registration error (TRE), in millimeters. The mean and the standard deviation of TRE for the four methods is reported in Table 1. We performed Wilcoxon rank-sum test on TRE distribution to compare the combined curve- and surface-based registration with the curve-based and surface-based methods individually. Results are reported in the Table 1. Box-plots in Fig. 2(a) show the overall accuracy of the four image registration methods on a complete set of landmarks over all five cases.

Correlation between TRE for the intensity-based and combined curve- and surface-based registration was $\rho = 0.5$, varying from $0.17 - 0.59$ for the five cases. Overall, for 35.5% cases of landmarks the combined curve- and surface-based registration method performed better than intensity-based method.

4 Discussion

Fig. 2(a) shows that the curve-based method alone provides good registration accuracy for the majority of landmarks. However, there are many outliers present with errors of up to 2.5 cm. Within the framework of currents, points located further than the typical scale of deformations σ_V are not affected by deformations of the features, which might cause landmarks distant to the vessel centerlines to be misaligned. Surface-based registration result in a slight overall improvement in TRE compare to the initial configuration. In contrast, incorporating both surfaces and curves into feature-based registration results in more accurate registration (1.85 mm) compared to both curve-based (2.37 mm) and surface-based (3.53 mm) methods.

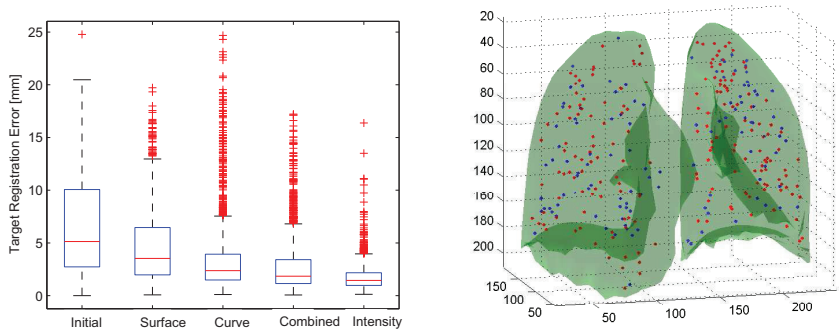
Table 1. Registration error at the landmark positions in [mm] for the four registration methods. The mean (m) and the standard deviation (sd) are reported. Statistical comparison of combined curve- and surface-based registration method was performed against the surface-based and curve-based methods. The notations of statistical significance level are as follows: * corresponds to $p < 0.05$ and ^{ns} to $p > 0.05$. The most right column indicates percentage of landmarks where the combined curve- and surface-based registration outperforms the intensity-based registration.

Image Registration Accuracy in mm [m ± sd]						
N	Before	Combined	Surface	Curve	Intensity	%
1	3.89 ± 2.78	1.47 ± 0.72	$2.45 \pm 1.56^*$	$2.24 \pm 1.41^*$	1.23 ± 0.61	37.7
2	4.34 ± 3.90	2.19 ± 1.98	$3.63 \pm 2.94^*$	2.32 ± 2.06^{ns}	1.26 ± 0.67	39.0
3	6.94 ± 4.05	3.30 ± 3.05	$5.31 \pm 3.26^*$	$3.03 \pm 2.79^*$	1.86 ± 1.11	25.0
4	9.83 ± 4.86	3.34 ± 2.67	$5.98 \pm 3.74^*$	$5.28 \pm 4.52^*$	2.15 ± 1.48	36.0
5	7.48 ± 5.51	3.83 ± 3.54	$5.80 \pm 4.37^*$	$4.40 \pm 4.42^*$	2.32 ± 1.82	40.0
All 5 cases						
	6.50 ± 4.83	2.83 ± 2.72	$4.63 \pm 3.58^*$	$3.45 \pm 3.48^*$	1.76 ± 1.31	35.5
median	5.13	1.85	3.53	2.37	1.44	

The median of TRE for the combined curve- and surface-based registration was 1.85 mm compared to 1.44 mm for the intensity-based method. Several reasons may lead to larger TRE for the combined curve- and surface-based method, such as inconsistency in segmentations of vessels in the two images. Ambiguous segmentation of lung surface near the hilum may leads to large missregistration errors in this area. Fig. 3(b) shows a difficult case in the data with irregular centerlines in the back of the lungs. Registration of lung images based on such geometrical structures as vessels centerlines and lung surfaces can be naturally improved by including airways and lung fissures into the presented framework.

In order to understand where are the main differences between the feature-based and intensity-based method, we visualized discrepancy between the two deformation fields in Fig. 3(a). For illustration purpose, we sparsely selected points where the orientation between deformation vectors were above 60° and with the magnitude of discrepancy vectors more than 3 mm and plotted inside the lung area. Interestingly, the discrepancy between the feature- and intensity-based methods were localized.

We further investigate image slices located at the areas where the discrepancy between the two methods was largest (blue cut planes in Fig. 3(a)). Fig. 4 shows the difference image with the moving image subtracted from the fixed image for both registration methods. Overall, lung surfaces and small vessels were aligned more accurately with the feature-based registration method.



(a) Box-plot of target registration errors. (b) Distribution of landmarks.

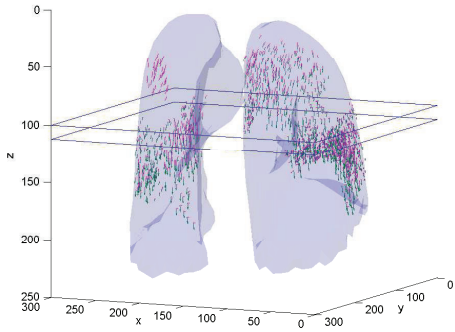
Fig. 2. Target registration errors (TRE) is shown in 2(a), as follows, before registration was applied (Initial), after surface-based (Suface), after curve-based (Curve), after combined curve- and surface-based (Combined) and after intensity-based registration (Intensity). Example 2(b) shows the spatial distribution of landmarks in the lungs. The landmarks, better aligned with the combined feature-based method are shown in red and with the intensity-based method in blue.

Another important component of currents is the length or the weight of the direction vector. For the task of registration of repeated lung CT images, the current for a small vessel could be given more weight than for a large vessel, leading to more accurate registration of small vessels. This is an important advantage of current-based registration over intensity-based method where small vessels with low contrast to surrounding lung tissue have negligible impact on the overall cost function. In this paper we used equal weights for all currents and normalized the length to 1.

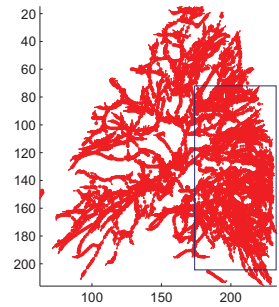
On average, 35.5% of landmarks were aligned better with the curve- and surface-based registration. The low correlation coefficient (0.5) suggests that the two registration methods align landmarks differently and may be combined into a more robust registration method.

5 Conclusion

In this paper, a curve- and surface-based registration method is presented, where lung surface and vessel tree centerlines are built-in into the framework of current-based registration. Incorporating both centerlines and surfaces results in more accurate registration than curve- or surface-based registration method alone. The proposed combined curve- and surface-based registration method achieves slightly lower accuracy than intensity-based registration but for 35.5% of landmarks outperformed the intensity-based method. A natural extension of the



(a) Deformation vectors for the combined curve- and surface-based (magenta) and intensity-based (green) methods methods



(b) Example of irregular current

Fig. 3. (a) An example of discrepancy in deformation fields between the feature-based and intensity-based registration methods. (b) An example of an ambigious current for the back of the lung.

presented work will be incorporating more anatomical lung structures, such as airways and fissures, to improve the feature-based method.

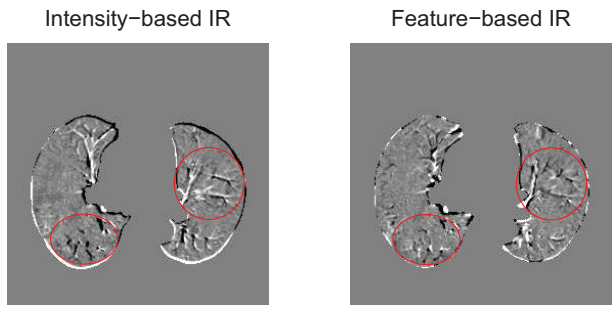
Results show that the proposed feature-based registration method is robust to inconsistent segmentation and outliers in segmented features and capable of handling imperfect segmentations. In applications where importance of different features varies, the prior weight of a feature may be encoded into the presented registration framework. Results suggest that a natural improvement of registration would be obtained by combining the feature- and intensity-based methods.

Acknowledgements. This work is financially supported by the Danish Council for Strategic Research under the Programme Commission for Nanoscience and Technology, Biotechnology and IT (NABIIT), the Netherlands Organization for Scientific Research (NWO), and AstraZeneca, Lund, Sweden.

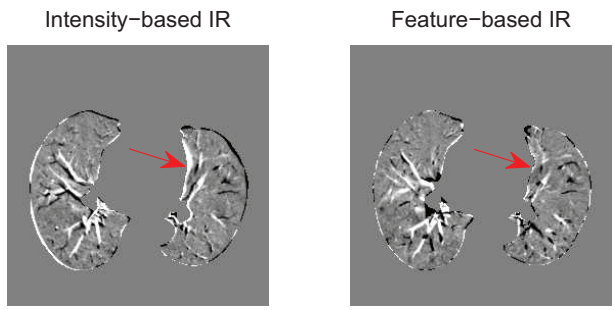
Authors would like to thank Juan Eugenio Iglesias, Medical Imaging Informatics UCLA, for useful suggestion.

References

1. Rohr, K., Stiehl, H.S., Sprengel, R., Buzug, T.M., Weese, J., Kuhn, M.H.: Landmark-based elastic registration using approximating thin-plate splines. IEEE Transactions on Medical Imaging **20**(6) (2001) 526–534
2. Johnson, H.J., Christensen, G.E.: Consistent landmark and intensity-based image registration. IEEE Transactions on Medical Imaging **21** (2002) 450–461



(a)



(b)

Fig. 4. Visual comparison of the combined feature-based and intensity-based registration methods. Slice cuts (a), (b) from the difference image between fixed and deformed image for the intensity- and combined feature-based registration methods were extracted on the same level as the plane cuts in Fig. 3(a). In general, the currents-based registration aligns the vessels and lung surface better, as can be seen in the areas indicated with the red circles and arrows.

3. Bookstein, F.L.: Morphological tools for landmark data; geometry and biology. Cambridge University press (1991)
4. Joshi, S.C., Miller, M.I.: Landmark matching via large deformation diffeomorphisms. IEEE Transactions on Image Processing **9** (2000) 1357–1370
5. Vaillant, M., Glaunes, J.: Surface matching via currents. Information Processing in Medical Imaging, 19th International Conference **3565** (2005) 381–392
6. Glaunès, J., Qiu, A., Miller, M.I., Younes, L.: Large deformation diffeomorphic metric curve mapping. Int. J. Comput. Vision **80**(3) (2008) 317–336
7. Durrleman, S., Pennec, X., Trouvé, A., Ayache, N.: Measuring brain variability via sulcal lines registration: a diffeomorphic approach. In: Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI), Part II. Volume 4791 of LNCS. (2007) 675–682
8. Vik, T., Kabus, S., von Berg, J., Ens, K., Dries, S., Klinder, T., Lorenz, C.: Validation and comparison of registration methods for free-breathing 4D lung-CT. In: SPIE Medical Imaging. (2008)
9. Li, P., Malsch, U., Bendl, R.: Combination of intensity-based image registration with 3D simulation in radiation therapy. Phys. Med. Biol **53** (2008) 4621–4637
10. Betke, M., Hong, H., Thomas, D., Prince, C., Ko, J.P.: Landmark detection in the chest and registration of lung surfaces with an application to nodule registration. Medical Image Analysis **7**(3) (2003) 265–281
11. Durrleman, S., Pennec, X., Trouvé, A., Ayache, N.: Sparse approximation of currents for statistics on curves and surfaces. In: Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI), Part II. Volume 5242 of LNCS. (2008) 390–398
12. Durrleman, S., Pennec, X., Trouvé, A., Thompson, P., Ayache, N.: Inferring brain variability from diffeomorphic deformations of currents: an integrative approach. Medical Image Analysis **12**(5) (2008) 626–637 PMID: 18658005.
13. Castillo, R., Castillo, E., Guerra, R., Johnson, V.E., McPhail, T., Garg, A.K., Guerrero, T.: A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. Physics in Medicine and Biology **54**(7) (2009) 1849–1870
14. Lo, P., Sporring, J., Ashraf, H., Pedersen, J., de Bruijne, M.: Vessel-guided airway segmentation based on voxel classification. In: Pulmonary Workshop at Medical Image Computing and Computer Assisted Intervention (MICCAI). (2008)
15. iso2mesh: A Matlab/Octave-based mesh generator. http://iso2mesh.sf.net
16. Wang, T., Basu, A.: A note on a fully parallel 3D thinning algorithm and its applications. Pattern Recognition Letters **28(4)** (2007) 501–506
17. Gorbunova, V., Lo, P., Ashraf, H., Dirksen, A., Nielsen, M., de Bruijne, M.: Weight preserving image registration for monitoring disease progression in lung CT. In: Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI), Part II. Volume 5242 of LNCS. (2008) 863–870

Towards Automated Detection of Abnormalities in Lung Segmentations

Reinhard Beichel[1,2,3], Simon Schlosser[4], Pierre Elbischger[4], Shanhui Sun[1,3], and Geoffrey McLennan[2,3]

[1] Dept. of Electrical and Computer Enginering
[2] Dept. of Internal Medicine
[3] The Iowa Institute for Biomedical Imaging
The University of Iowa, Iowa City, IA 52242, USA
reinhard-beichel@uiowa.edu

[4] Dept. of Medical Information Technology
Carinthia University of Applied Sciences
A-9020 Klagenfurt, Austria

Abstract. Automated lung segmentation in multidetector computed tomography data is a first processing step in computer-aided quantitative assessment of lung disease. Robust segmentation of diseased lungs is a non-trivial problem which is unsolved up to now. Consequently, lung segmentation results need to be manually verified, which is time-consuming and costly. We propose a novel algorithm for detecting gross abnormal lung segmentations based on a fast 3D shape retrieval approach. First, the segmentation result to verify is used to query a 3D lung shape database containing normal lung shapes. Second, the 3D shape dissimilarity between query and retrieved shape is utilized to assess the abnormality of the segmentation. Our method represents a first step toward the development of a quality assessment system for lung segmentations.

Key words: Segmentation abnormality detection, shape retrieval, shape context, lung segmentation

1 Introduction

Lung diseases like cancer, chronic obstructive pulmonary disease (COPD) or pneumonia are a major health problem. Multidetector computed tomography (MDCT) based lung imaging plays an important role in the early detection, diagnosis, and treatment of lung disease. Automated image analysis of lung MDCT data supports physicians in the quantitative assessment of lung disease [1]. One of the first steps in lung image analysis is to segment the lungs. Several approaches to lung segmentation in MDCT have been proposed. An overview can be found in [1]. The majority of approaches are based on gray-value analysis and assume that there is a large density difference between lung tissue and surrounding structures. In case of normal lungs, such approaches produce sufficiently accurate results. However, they frequently fail to deliver correct lung segmentations

if the lungs contain high density pathology regions like masses (tumors) or pneumonia. In general, the segmentation of diseased lungs is a non-trivial problem which is important for clinical applications and research.

Recently, a few papers proposing more robust lung segmentation algorithms have been published. For example, Sluimer et al. propose a segmentation by registration scheme to robustly segment diseased lungs [2, 3]. While delivering promising results, not all pathological cases could be handled successfully [3]. In addition, the proposed approach is quite computing-time intensive [3]. An adaptive border marching algorithm was presented in [4] to reliably include juxtapleural nodules in lung segmentations. An approach for the robust segmentation of lung parenchyma based on the curvature of ribs was presented in [5]. In summary, up to now available robust lung segmentation approaches are not able to handle all possible pathological cases successfully and/or require significantly more computing-time than conventional lung segmentation methods. This is especially a problem for the automated analysis of large numbers of lung MDCT data sets, as required in multi-center clinical trials or the automated processing/analysis of lung image databases collected by the Lung Image Database Consortium (LIDC) [6] or the Reference Image Database to Evaluate Response (RIDER)[1] projects, for example. Even the verification of the correctness of several thousand lung segmentation results is a costly and time-consuming task. To efficiently deal with this problem, an automated quality assessment/control system is needed that allows the automatic identification of abnormal lung segmentations (e.g., segmentation errors, breathing artifacts, etc.). Once abnormal segmentations are identified, different remedy strategies can be applied, including:

- adapt parameters and rerun the segmentation algorithm,
- apply a more robust—but usually more time consuming—method,
- manually correct the automatically generated segmentation result,
- switch to a semiautomatic/manual segmentation approach, or
- exclude the data set from further analysis.

A process might even utilize a sequence of the above outlined strategies depending on the classification result of the segmentation after each processing step.

In this paper we present a novel approach to detect grossly abnormal lung segmentations based on a shape retrieval method. The main idea behind our approach is as follows. First, the 3D lung segmentation result is used to query a reference lung shape database. Data sets in the reference database are representatives for normal lung shapes. Second, an assessment of the segmentation abnormality is derived from a 3D shape dissimilarity between query shape and the best match found in the reference database. Our approach builds on a fast shape retrieval approach based on shape contexts which we recently introduced for 2D shape retrieval [7]. This allows the rapid assessment of segmentation results. While content-based image retrieval systems have been developed for lung images [8, 9], we are not aware of similar approaches for 3D lung shape.

[1] http://ncia.nci.nih.gov/collections

Abnormalities in 3D shape can be caused by segmentation errors (e.g., due to high density lung pathology), breathing artifacts, pleural effusion, missing data and so on. The work described in this paper represents a first step toward the development of a lung segmentation quality assessment system that can differentiate between different causes of abnormal segmentations.

2 Related Work

Content-based retrieval systems are a prerequisite to effectively utilize databases like medical image archives. The two most popular query techniques in this context are semantic retrieval and query by example. In the latter case, the user provides an example depicting the content of the query. To process such an request, algorithms are needed that assess the similarity/dissimilarity between query and examples stored in a database. This allows for the ordering of the database examples based upon their relevance, and to retrieve the closest match.

Shape is an important feature for querying object databases. Several methods for 3D shape retrieval have been proposed. A review of methods can be found in [10]. In the case of large databases, the time needed to assess the similarity of images is critical for retrieval performance, and thus, for practical applicability.

Shape contexts [11, 12] represent a powerful method for shape description and similarity assessment. They have been used for applications like shape retrieval [13, 14] or hippocampal surface mapping [15]. Fig. 1 illustrates the basic idea of shape contexts on 2D shapes. Given an object in an image, a representation

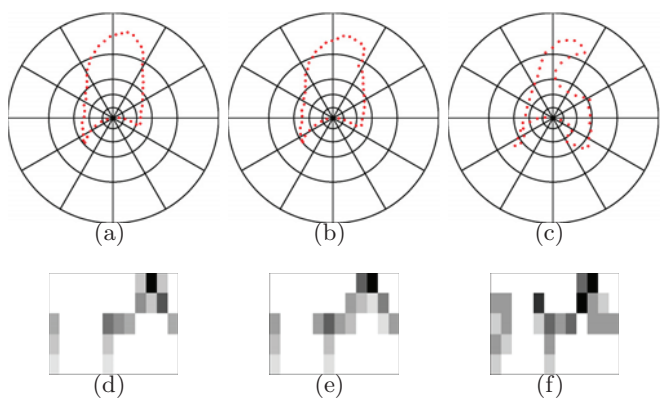


Fig. 1. Examples of 2D shape contexts. (a-c) Sampled object shapes shown with log-polar histogram bins placed on a reference point. (d-f) Shape contexts corresponding to (a-c).

of the object's shape based on a finite set of 2D points $P = \{\mathbf{p}_1, ..., \mathbf{p}_n\}$ with $\mathbf{p}_i \in \mathbb{R}^2$ is generated by sampling the object contour. Point sampling can be

done randomly or with an equidistant spacing. A reasonable selection of n, the number of sample points, is required to have a good representation of the input shape. For a point $\mathbf{p}_i$ of the shape represented by P, a coarse histogram $\mathbf{H}_i$ of the relative coordinates of the remaining $n-1$ points is computed:

$$\mathbf{H}_i(k) = \#\{\mathbf{q} \neq \mathbf{p} | (\mathbf{q} - \mathbf{p}_i) \in bin(k)\} \; . \tag{1}$$

The calculated histogram $\mathbf{H}_i$ is denoted as the shape context of $\mathbf{p}_i$ [12]. For example, Figs. 1(a-c) depict three different 2D shapes, each with a selected reference point $\mathbf{p}_i$ and a log-polar histogram outline around $\mathbf{p}_i$. The corresponding shape contexts are depicted in Fig. 1(d-f). The shape context in Figs. 1(d) and 1(e) show similarities, because the object shapes are similar (normal lung shapes) and reference points on both of the shapes have been chosen such that they are approximately corresponding. In contrast, the similarities of the shape contexts of the normal lung shapes (Figs. 1(d-e)) and the abnormal lung shape (Fig. 1(f)) are quite low. Consequently, shape contexts can be utilized to compare shapes.

The bins of the histogram are usually chosen such that they are uniform in log-polar space, which makes the descriptor more sensitive to shape points nearby the point $\mathbf{p}_i$. Normalization to rotation of a shape is achieved by aligning the histogram with the shape tangent or normal vector direction in point $\mathbf{p}_i$. Invariance to isotropic scaling can be gained by normalizing the shapes regarding their size before they are used to calculate the shape context.

For shape retrieval, the similarity/dissimilarity between the query shape and all the shapes in a database need to be assessed. For example, this can be accomplished by using shape contexts. As outlined in [7], a major drawback of utilizing shape contexts directly for shape retrieval is the high computing-time needed for the comparison of a query shape with shapes in a database, since each shape is usually described by several hundred or thousand shape contexts which need to be compared individually. Consequently, methods that utilize a single shape signature are preferable for shape retrieval, especially when large databases are queried.

3 Method

Our approach to detect gross abnormalities in lung segmentations consists of two main processing steps. First, the given lung segmentation is used to query a 3D lung shape database containing normal lung shapes. Second, the 3D shape dissimilarity between retrieval result and query segmentation is used to assess the segmentation. The left and right lung are processed independently, thus we utilize two databases of normal left and right lung shapes. However, the processing scheme is the same for left and right lungs. The individual components of our method are described in detail in the following sections.

3.1 Fast 3D Shape Retrieval

For efficient 3D shape retrieval from a database, we use a shape signature which is based on shape contexts. The utilized approach represents an extension of

our previous work on fast 2D shape retrieval [7] to 3D. The main idea behind our approach is outlined in Fig. 3. For a given 3D shape, represented by a point set $P = \{\mathbf{p}_1, \ldots, \mathbf{p}_o\}$ with $\mathbf{p}_i \in \mathbb{R}^3$ (Figs. 3(a-c)), all the corresponding shape histograms $\mathbf{H}_i$ with $i = 1, \ldots, o$ are calculated. The 3D (spherical) shape contexts $\mathbf{H}_i$ (Fig. 2) are then rearranged into 1D shape context vectors $\mathbf{v}_i$, based on which a single shape signature is calculated for P (Figs. 3(d-f)). The similarity of two shapes can then be assessed by comparing the corresponding shape signatures. Basically, the task of assessing shape similarity is transformed into the task of assessing the similarity of shape signatures. Consequently, in terms of computing time, expensive point-to-point shape context comparisons are avoided.

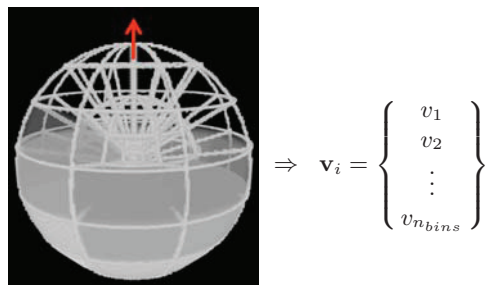


Fig. 2. Example of a 3D (spherical) shape context histogram $\mathbf{H_i}$ and the corresponding rearranged shape context vector $\mathbf{v_i}$. A log-polar partitioning scheme is used to define the 3D histogram bins. To gain invariance to rotation, the axis of the shape histogram (red arrow) is aligned with the surface normal at a sample point $\mathbf{p}_i$. The bin content is rendered in gray.

Given two 3D point sets $P_x = \{\mathbf{p}_{x,1}, \ldots, \mathbf{p}_{x,m}\}$ and $P_y = \{\mathbf{p}_{y,1}, \ldots, \mathbf{p}_{y,n}\}$ to compare, let the matrices $\mathbf{S}_x = [\mathbf{v}_{x,1}, \ldots, \mathbf{v}_{x,m}]^T$ and $\mathbf{S}_y = [\mathbf{v}_{y,1}, \ldots, \mathbf{v}_{y,n}]^T$ store all the rearranged shape context vectors of the shapes P_x and P_y, respectively. We then define the following shape dissimilarity measure:

$$SD(P_x, P_y) = 1 - corr(scc(P_x), scc(P_y)) \,, \tag{2}$$

where the shape context covariance (signature) of a point set P is defined as $scc(P) = cov(\mathbf{S})$, and $cov(\mathbf{S})$ denotes the covariance matrix of all the rows (shape context vectors) in $\mathbf{S}$. The function $corr$ in Eq. 2 denotes the normalized correlation coefficient. Low values for SD indicate a high similarity, and large values indicate shapes with low similarity. Given a query shape P_{query} and a normal lung shape database $DB = \{P_1, \ldots, P_w\}$, the shape retrieval result $P_{i_{match}}$ is generated by evaluating $i_{match} = \arg\min_{i=1,\ldots,w}\{SD(P_{query}, P_i)\}$ with $P_i \in DB$. To speedup the retrieval process, all shape signatures $scc(P_i)$ of database shapes can be calculated in an offline processing step and stored in the database. In addition, only the upper triangular part of $scc(P)$ needs to be

considered for correlation calculation and storage, because of the symmetry of the covariance matrix. This reduces computing-time and the required storage space for the shape signatures. Matrices $\mathbf{S}_x$ and $\mathbf{S}_y$ are of size $m \times n_{bins}$ and $n \times n_{bins}$, respectively. n_{bins} denotes the number of histogram bins. The numbers m and n of points in each point set can be different, but we prefer that m and n are the same. Consequently, we randomly sample $u = m = n$ points from the surface of the left and right segmented lung shape.

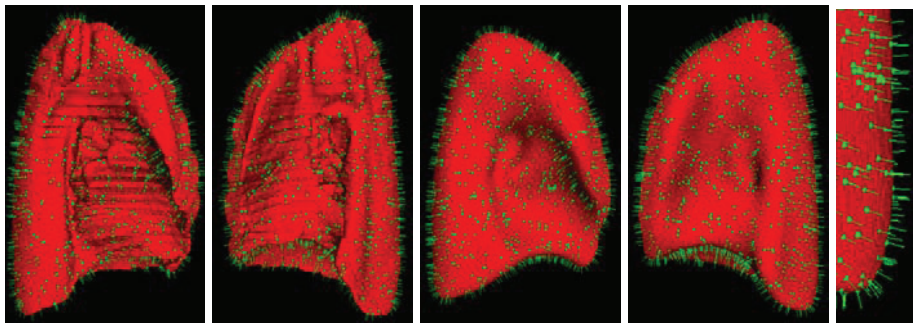
Fig. 3. 2D example of shape signatures derived from shape contexts. (a-c) Point sets representing shapes. (d-f) Corresponding shape signatures $\mathbf{H}$; a nonlinear gray-value transformation was applied for better visualization of signature patterns.

To gain invariance to rotation, a surface normal vector is calculated for each sample point (Fig. 4(e)) to align the shape context. To robustly compensate for differences in size, we utilize the following approach. The left and right lungs of the database are roughly normalized before the shape signature is calculated by isotropically scaling them so that they have a height of $h_l = 260$ mm (approx. average lung size in z-direction). To calculate shape contexts of database shapes, a minimum radius of $r_{min_{DB}} = 6$ mm and maximum radius of $r_{max_{DB}} = h_l/3$ is used. Furthermore, after sampling shape points from the query data set, several shape signatures $scc(P_{query,\gamma}) = scc_\gamma(P_{query})$ are calculated by using $r_{min_{Q\gamma}} = \gamma r_{min_{DB}}$ and $r_{max_{Q\gamma}} = \gamma r_{max_{DB}}$ with the scale factor $\gamma \in O$ and $O = \{0.5, 0.6, \ldots, 1.5\}$ for the log-polar histograms. The size invariant shape retrieval match $P_{i_{match\_SI}}$ is found by evaluating

$$i_{match\_SI} = \arg\min_{i=1,\ldots,w} \{SD(P_{query,\gamma_i}, P_i)\} \tag{3}$$

with $\gamma_i = \arg\min_{\gamma \in O}\{SD(P_{query,\gamma}, P_i)\}$. The corresponding parameter $\gamma_{i_{match\_SI}}$ represents a discrete estimate for the relative size between query shape and retrieved normalized shape from the lung database.

Depending on the imaging protocol used, the scans can show heart motion artifacts. In addition, ribs, airways, as well as pulmonary arteries and veins can cause certain local variations in shape (Figs. 4(a) and (b)). To better deal with this variation, we smooth the volumetric lung shape data sets by applying a convolution with a Gaussian kernel ($\sigma = 10$) before the surface points are sampled. Examples for smoothed lung shapes are shown in Figs. 4(c) and (d). In comparison, the corresponding original shapes are shown in Figs. 4(a) and (b).



(a) (b) (c) (d) (e)

Fig. 4. Sampled surface points utilized to describe lung shapes shown with corresponding normal vectors. (a) Left and (b) right lung. (c) Smoothed lung shape corresponding to (a). (d) Smoothed lung shape corresponding to (b). (e) Magnified lower right section of the smoothed lung shown in (d).

Let $SD_{DB}(P_{query})$ denote the dissimilarity between query and retrieved database shape. Gross shape abnormalities can be detected by applying a threshold to $SD_{DB}(P_{query})$.

4 Experiments

For our experiments, we utilized 54 left and 57 right lung shapes for building a normal lung shape database. In addition, 22 left and 16 right gross abnormal lung segmentations were available. Abnormalities were caused by different kinds of lung pathology (e.g., cancer), missing data, segmentation errors, and severe breathing artifacts. All lung shapes were represented by $u = 4000$ randomly selected surface sample points. The selection of u represents a good trade-off between computing time, needed for calculating shape histograms, and the ability to describe local surface details. Shape histograms with $n_{bins} = 360$ were utilized (partitioning scheme: 5, 8, and 9 bins for radius, longitude, and latitude, respectively; see Fig. 2).

Two experiments were performed to assess our method. First, we performed a leave-one-out experiment on the database shapes. There are two reasons behind this experiment. First, the experiment allows us to find out how well the left out shape is represented in the database. This is an important analysis, because we have a limited number of database shapes for the left and right lung. Clearly, the more lung shape cases are stored in the database, the better the population can be described. Second, it allows us to find out which 3D shape dissimilarity (SD) values we can expect for normal lung shapes. In the second experiment, the available abnormal left and right lung shapes were used to query the corresponding normal lung shape databases. The SD histograms of both experiments are plotted in Figs. 5(a) and (b) for the left and right lung, respectively. And the corresponding receiver operating characteristics (ROC) are shown in Fig. 6.

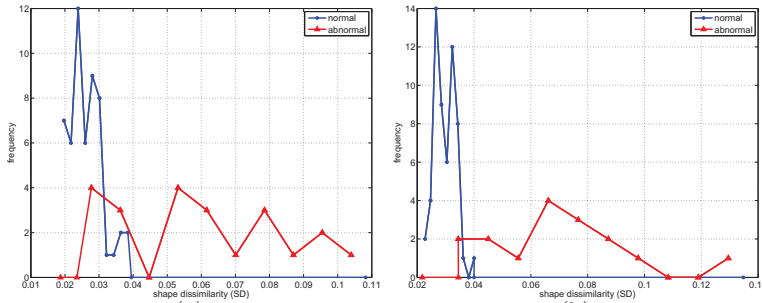


Fig. 5. Shape dissimilarity (SD) histograms for normal and gross abnormal shapes. (a) Left lung and (b) right lung.

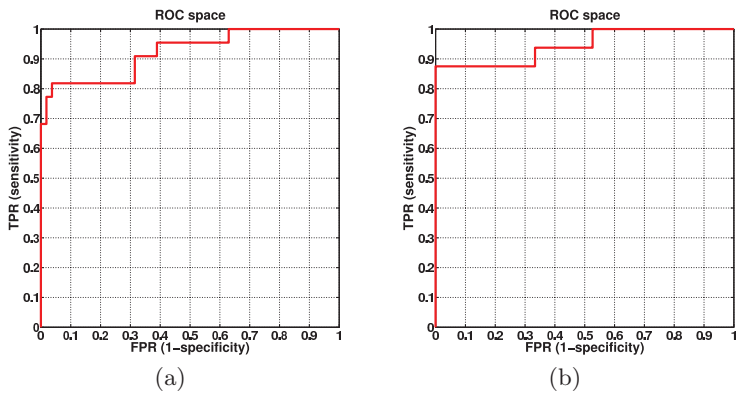
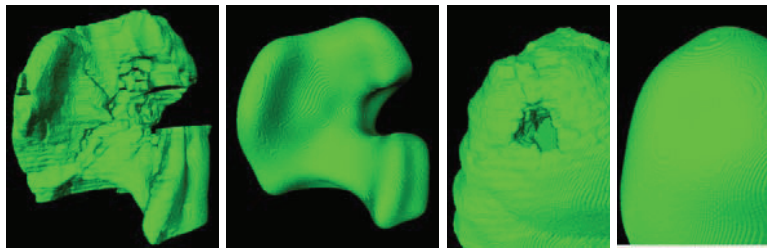


Fig. 6. Receiver operating characteristics (ROC) for the (a) left and (b) right lung.

The time required for producing a point representation of a query segmentation was 220 seconds. The generation of the eleven shape signatures needed for gaining size invariance took 72 seconds. The intrinsic shape retrieval process required 0.5 seconds on average. Our experiments were performed on a standard image pressing workstation, and the utilized implementation was not optimized for speed.

5 Discussion

The shape dissimilarity (SD) histograms in Figs. 5 show that normal lung shapes have quite low SD values compared to abnormal shapes. For the left and right lung shapes, there is some overlap in terms of shape dissimilarity. This is also clearly visible in the ROC plots in Fig. 6. One cause for this might be that some normal shapes are not well represented in our database. For the reported experiments, a limited number of lung shapes was available. Clearly, the currently



(a) (b) (c) (d)

Fig. 7. Impact of surface smoothing. (a) and (c) Lung segmentations containing segmentation errors; (b) and (d) corresponding smoothed lung shapes. The gross segmentation error is clearly visible in (b), but the error caused by a tumor in (c) is not visible in (d) due to size.

utilized number of database shapes is not representative for all possible lung shapes, as indicated by our experiments. For example, barrel chest "shaped" lungs are underrepresented. In this context, we plan to evaluate our method on a larger data collection in the future.

Using smoothed lung shapes for analysis clearly represents a trade-off. On the one hand, only basic shape characteristics are utilized for shape analysis, and heart motion artifacts, rib patterns on the lung surface, etc. do not influence the shape dissimilarity measure. On the other hand, some (smaller) abnormalities might not be detected, because they are not or only allusively represented by the smoothed shape (Fig. 7). In this paper, we are specifically aiming at assessing/detecting gross abnormalities based on shape analysis. In the future we plan to expand our approach by utilizing shape representations on different scales. This will allow us to detect more and smaller abnormalities. In this context, calculating shape signatures for specific sub-parts of lung shapes might also enable to detect smaller and local abnormalities.

6 Conclusion

We have presented an approach to assess grossly abnormal lung segmentations based on a shape retrieval approach. Our method represents a step towards the development of an automated quality assessment system. Such a system will be beneficial for the automated analysis of large numbers of MDCT lung scans. In addition, it offers the prospect of developing novel approaches to robust lung segmentation.

7 Acknowledgments

The authors thank Sudharshan Reddy Bommu and Joseph Reinhardt at The University of Iowa for providing segmented lung data sets. This work was supported in part by the Austrian Marshall Plan Foundation.

References

1. Sluimer, I., Schilham, A., Prokop, M., van Ginneken, B.: Computer analysis of computed tomography scans of the lung: a survey. IEEE Trans. Med. Imag. **25**(4) (2006) 385–405
2. Sluimer, I.C., Niemeijer, M., van Ginneken, B.: Lung field segmentation from thin-slice CT scans in presence of severe pathology. In: Proc. SPIE (Medical Imaging). Volume 5370. (2004) 1447–1455
3. Sluimer, I., Prokop, M., van Ginneken, B.: Toward automated segmentation of the pathological lung in CT. IEEE Trans. Med. Imag. **24**(8) (2005) 1025–1038
4. Pu, J., Roos, J., Yi, C.A., Napel, S., Rubin, G.D., Paik, D.S.: Adaptive border marching algorithm: Automatic lung segmentation on chest CT images. Computerized Medical Imaging and Graphics **32**(6) (2008) 452 – 462
5. Prasad, M.N., Brown, M.S., Ahmad, S., Abtin, F., Allen, J., da Costa, I., Kim, H.J., McNitt-Gray, M.F., Goldin, J.G.: Automatic segmentation of lung parenchyma in the presence of diseases based on curvature of ribs. Academic Radiology **15**(9) (2008) 1173 – 1180
6. Meyer, C.R., Johnson, T.D., McLennan, G., Aberle, D.R., Kazerooni, E.A., Macmahon, H., Mullan, B.F., Yankelevitz, D.F., van Beek, E.J.R., Armato, S.G., McNitt-Gray, M.F., Reeves, A.P., Gur, D., Henschke, C.I., Hoffman, E.A., Bland, P.H., Laderach, G., Pais, R., Qing, D., Piker, C., Guo, J., Starkey, A., Max, D., Croft, B.Y., Clarke, L.P.: Evaluation of lung MDCT nodule annotation across radiologists and methods. Acad Radiol **13**(10) (Oct 2006) 1254–1265
7. Schlosser, S., Beichel, R.: Fast shape retrieval based on shape contexts. In: Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis (ISPA), Salzburg, Austria (September 2009) in print.
8. Dy, J.G., Brodley, C.E., Kak, A., Broderick, L.S., Aisen, A.M.: Unsupervised feature selection applied to content-based retrieval of lung images. IEEE Trans. Pattern Anal. Mach. Intell. **25**(3) (2003) 373–378
9. Aisen, A.M., Broderick, L.S., Winer-Muram, H., Brodley, C.E., Kak, A.C., Pavlopoulou, C., Dy, J., Shyu, C.R., Marchiori, A.: Automated Storage and Retrieval of Thin-Section CT Images to Assist Diagnosis: System Description and Preliminary Assessment. Radiology **228**(1) (2003) 265–270
10. Tangelder, J.W.H., Veltkamp, R.C.: A survey of content based 3D shape retrieval methods. Journal Multimedia Tools and Applications **39**(3) (2008) 441–471
11. Belongie, S., Malik, J.: Matching with shape contexts. In: Proc. IEEE Workshop on Content-based Access of Image and Video Libraries. (2000) 20–26
12. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence **24** (2002) 509–522
13. Daliri, M.R., Torre, V.: Robust symbolic representation for shape recognition and retrieval. Pattern Recogn. **41**(5) (2008) 1782–1798
14. Leibe, B., Schiele, B.: Analyzing appearance and contour based methods for object categorization. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Volume 2. (2003) 409–415
15. Shi, Y., Thompson, P.M., de Zubicaray, G.I., Rose, S.E., Tu, Z., Dinov, I., Toga, A.W.: Direct mapping of hippocampal surfaces with intrinsic shape context. NeuroImage **37**(3) (2007) 792–807

Obstructive pulmonary function: Patient classification using 3D registration of inspiration and expiration CT images

K. Murphy, B. van Ginneken, E.M van Rikxoort, B.J. de Hoop, M. Prokop, P. Lo, M. de Bruijne, and J.P.W. Pluim

University Medical Center, Utrecht, The Netherlands.

Abstract. Chronic Obstructive Pulmonary Disorder (COPD) is a condition with 4 classes of severity, characterised by reduced airflow and diagnosed by means of pulmonary function testing. CT scanning offers far more detailed information about the underlying pathology and site(s) of the disease than is afforded by a pulmonary function test, but is not currently used in the diagnosis and classification of COPD. In this work a classification system is presented to classify 110 subjects in a COPD database based on 2 thoracic CT scans taken at full inspiration and at full expiration for each subject. Experiments with a 2-class kNN classifier (COPD/non-COPD) as well as with a 5-class kNN classifier (COPD 1-4/non-COPD) are presented. Features are derived from 3 sources 1)The inspiration scan, 2)The expiration scan, 3)Comparison of the HU values at inspiration and expiration based on a 3D, fully automatic, non-rigid registration.
Classification in a 2-class system is particularly successful achieving an area under the ROC curve of Az=0.92. Multi-class classification is more challenging, with 66% of cases correctly classified and a further 28% classified in a class neighbouring the true one. A discussion is presented of the difficulties of precise multi-class classification due to drawbacks of the reference standard of pulmonary function testing. It is shown that the inclusion of registration-derived features improved the performance of the classifiers in all cases.

1 Introduction

Chronic Obstructive Pulmonary Disorder (COPD) is the fourth highest cause of chronic morbidity and mortality in the United States and by 2020 it is expected to be the fifth most important illness worldwide in terms of burden of disease [10]. COPD is characterised by airflow limitation which is not fully reversible and diagnosis is confirmed by means of spirometry (pulmonary function testing) [10]. Four levels of severity are defined according to the spirometry results (COPD classes 1-4), and the status of subjects without COPD is here referred to as 'class 0'. There may be several underlying causes of the airflow obstruction including emphysema and/or narrowing of the airways. While pulmonary function tests provide useful information about the overall extent of airflow restriction they do

not infer the pathological causes of this restriction or the regions of the lungs affected.

CT scans are used to examine the physical state of the lungs in COPD patients and to investigate the causes of the disease in more detail. CT-based measurements can be expected to provide quantitative information about airflow and have been shown to do so in many studies including [5,9]. In [9], for example, analysis of CT scans is used to compare the relative contributions of emphysema and airway-narrowing. Using measurements based on CT values creates the potential to simultaneously evaluate the individual components of the disease and its location within the lungs thereby improving physician's understanding of the mechanisms of the disease [11].

Both inspiration and expiration scans have a role to play in airflow analysis since emphysema is most easily visualised in inspiration scans, while regions of air-trapping (which may be caused by airway obstruction) are more obvious at full expiration. A direct regional comparison of CT scans taken at full inspiration and expiration is expected to provide valuable additional information, but visual comparisons are typically awkward to carry out due to the difficulty of aligning small regions in both scans. For a thorough quantitative analysis a full 3D registration of the scans is required.

In this work we attempt to determine whether quantitative airflow analysis by means of CT alone is sufficient firstly to diagnose COPD and secondly to categorise the subject into the correct COPD class using a k-Nearest-Neighbour (kNN) [3] classifier. Experiments with COPD diagnosis (COPD/non-COPD) were carried out using a 2-class kNN classifier. A 5-class kNN classifier was utilised in separate experiments to attempt to classify subjects into the correct COPD class (0-4). Features are extracted from both the inspiration scan (emphysema) and the expiration scan (air-trapping). Furthermore, we perform a fully automatic 3D registration of the inspiration and expiration scans, and extract additional features from a voxel by voxel comparison of the inspiration and expiration intensities. The various feature sets are compared to determine which features are most important in achieving an accurate classification.

2 Materials

In this study 110 pairs of scans are used which form part of a COPD database being constructed at our research institute. The subjects are either clinical patients or were recruited as part of a CT screening trial. All subjects are either at risk of developing COPD (former/current heavy smokers) or already have COPD. The numbers of scan pairs representing each COPD level (0-4, see section 3) are as follows: 0:29, 1:19, 2:33, 3:16, 4:13.

Each subject underwent a breath-hold CT scan at full inspiration and another at full expiration. All scans were acquired in less than 12 seconds. CT scanning protocol varied slightly depending on the subject but was usually set with a beam current of 30mAs for inspiration scans (low dose) and 20mAs for expiration scans (ultra-low dose).

All scans were obtained without contrast injection on a 16 detector-row scanner (Mx8000 IDT or Brilliance 16P, Philips Medical Systems). They have a per-slice resolution of 512×512, with the number of slices per scan varying per subject. Slice thickness is 1mm with overlapping slices and a slice-spacing of 0.7mm

3 Reference Standard

COPD is diagnosed by means of spirometry (pulmonary function testing). During pulmonary function tests subjects are instructed to exhale fully into a mouthpiece and various measurements are made, including 'Forced Expiratory Volume in 1 Second' (FEV1) and 'Forced Vital Capacity' (FVC). FEV1 measures how much air volume is released in the first second of expiration, and FVC determines the entire volume exhaled. A value known as 'FEV1-Predicted' is determined using look up tables based on details such as the subject's gender, height, weight, age and race. Table 1 illustrates how COPD is diagnosed based on these measurements. FEV1 as a percentage of FVC determines the presence or absence of COPD, while FEV1 as a percentage of FEV1-Predicted is used to establish the severity of the condition.

COPD Class:	0 (No COPD)	1 (Mild)	2 (Moderate)	3 (Severe)	4 (V. Severe)
Measurement:					
FEV1 as % of FVC	≥70	<70	<70	<70	<70
FEV1 as % of FEV1 Predicted		≥ 80	≥ 50, <80	≥ 30, <50	<30

Table 1. The criteria for determination of COPD class by pulmonary function testing

4 Methods

4.1 Pre-Processing

As an initial pre-processing step all inspiration and expiration scans were subsampled in order to make certain subsequent procedures more efficient. The images were reduced by block averaging to 256×256 voxels in the X-Y plane with the number of slices reduced such that the data was isotropically sampled. Linear interpolation was used to determine grey-values between voxel locations.

Segmentation masks for the lungs and lobes were computed for all scans. The lung segmentation algorithm was by Sluimer et al. [15] based on the method of Hu et al. [6]. Lobe segmentation was performed using an algorithm by van Rikxoort et al. [17]. All segmentations included in this study were visually checked.

Distance transforms from the lung boundaries were computed as follows: The lung mask was eroded with a spherical kernel of radius 5 voxels and a distance transform was calculated on the eroded segmentation. The erosion served to ensure that voxels inside the true lung boundary also obtained a distance value in the transform. Values above 10 in the distance transform image were clamped. These distance transform images were subsequently used in the registration procedure.

4.2 Registration

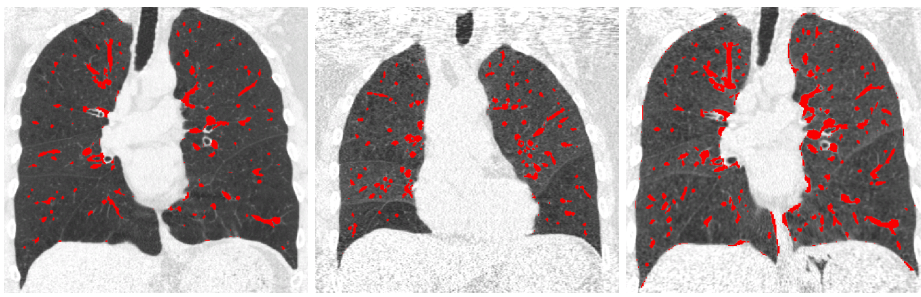
Registration was carried out on the down-sampled images in order to reduce memory consumption. The registration was implemented using `elastix` version 4.0 (http://elastix.isi.uu.nl) which is a registration toolkit based on the National Library of Medicine Insight Segmentation and Registration Toolkit (ITK). The expiration scan was transformed to match the inspiration scan and the calculated transform was subsequently applied to the original full resolution expiration data. The registration procedure consisted of an initial affine registration step followed by a non-rigid registration to handle the deformations of the lung tissue. Both steps involved a multi-resolution strategy with 5 resolution levels for the affine procedure and 7 for the non-rigid. The cost function used was a combination of the mutual information (MI) [16] of the down-sampled scans and the sum of squared differences (SSD) of the lung boundary distance images described in section 4.1. The lung boundary distance images were included as an additional guide to the registration to ensure that the lung boundaries were well aligned. The MI cost was weighted more heavily (0.75) than the SSD cost (0.25). Optimisation was by means of a stochastic gradient descent optimizer [7]. The non-rigid registration deformations were modelled by a B-Spline grid [12]. The grid-size varied per resolution-level with the finest grid at the last level having a spacing of 10mm in each dimension. The lung segmentation described in section 4.1 was used as a mask for the inspiration scan to ensure that during registration the cost function was calculated on samples from the volume of the segmented lungs only. The registration settings were chosen bearing in mind that there may be very large deformations between the inspiration and expiration scans. Registration using these settings took approximately 45 minutes per scan pair.

4.3 Vessel Segmentation

In order to exclude vessels from subsequent feature calculations (in which only lung parenchyma is of interest) a vessel segmentation is performed. Although the registration was reasonably accurate for most vessel structures, on some occassions vessel aligment was imperfect. For this reason vessel segmentation was carried out on the inspiration and transformed expiration images separately. Vessel segmentation was also carried out on the original expiration image.

The vessels were segmented using the algorithm of Lo et. al. [8]. Since the majority of the scans used a low-dose or an ultra-low dose protocol, the best results were obtained by using the full resolution scan data and applying a noise

filter [13] prior to the segmentation. The vessel segmentation was dilated with a spherical kernel of radius 1 voxel to ensure that undersegmentation was not an issue. In addition a conservative thresholding was applied to the original image to segment the brightest voxels. The final segmentation comprised the union of the dilated vessel segmentation and the thresholded bright voxels. The threshold t was set at $t = CT_{min} + ((CT_{max} - CT_{min}) \times 0.5)$ where CT_{min} and CT_{max} represent the minimum and maximum intensity values within the lung volume. In practise, to avoid CT_{min} and CT_{max} being influenced by noise or minor segmentation errors, CT_{min} is given the value of the 1st centile and CT_{max} the value of the 99th centile. Examples of vessel segmentations in inspiration and expiration are shown in figure 1. Note that there are occasionally voxels on the pleural surface which are included in the vessel segmentation due to the thresholding step and inaccuracies in the lung segmentation. These are not problematic in our application as our vessel segmentation serves only to exclude these voxels during feature calculation.



(a) (b) (c)

Fig. 1. Vessel Segmentations in (a) inspiration, (b) expiration and (c) transformed expiration

4.4 Feature Calculation

With the aim of classifying the subjects based on their COPD severity a number of features were calculated for each image pair. The features can be divided into 3 subsets as follows: Feature set A - Features of the inspiration image only, Feature set B - Features of the expiration image only, Feature set C - Features based on a voxel-by-voxel comparison of the inspiration image with the registered (transformed) expiration image. The features consist of percentages and averages which are calculated over both lungs, over each lung individually and over each lobe individually. Per-lung and per-lobe features are included since they may prove more useful than global features in the classification of COPD. The full list of 48 calculated features is given in table 2.

Feature calculation was carried out on the down-sampled versions of the images to conserve memory usage. Prior to calculating the features the inspiration, expiration and transformed expiration scan were noise-filtered [13]. Blood mass correction (BMC) [4] was applied to the transformed expiration image before the calculation of feature set C where the inspiration and transformed expiration images are compared directly. The theory of BMC is based on the assumption that the total mass of the lungs is not changed by the actions of inhaling and exhaling air. Calculating the masses according to the CT intensity values however, it can be demonstrated that a minor change in mass occurs. This change in mass is due to alterations in blood-flow in the lungs during the breathing process [2]. The BMC calculates the difference in mass for each lung between inspiration and expiration and thereby derives a correction factor for voxels in that lung. The grey-value of each voxel in the lung is adjusted in the transformed expiration image such that the corrected mass of the lung is equal to that of the same lung in inspiration.

Comparison of subjects based on the relationship between their inspiration and expiration scans can cause difficulties. For example, subject A may show a smaller change in intensity values between inspiration and expiration then subject B, which might suggest that subject A has a comparatively reduced airflow. However such a difference may also be due to the fact that subject A did not inhale or exhale as completely or deeply as subject B for the purposes of the CT scans. For this reason, all features in feature set C were normalised by multiplication with the value C calculated by $C = (Vol_{INSP} - Vol_{EXP})/Vol_{INSP}$ where Vol_{INSP} is the total lung volume at inspiration and Vol_{EXP} the total lung volume at expiration. Vessel exclusion during calculation of feature set C was achieved by excluding any voxel which was segmented as vessel in either the inspiration image or the transformed expiration image.

ID	Description	Feature set
1	Emphysema(-950). Percentage of voxels below -950HU in insp.	A
2	Emphysema(-910). Percentage of voxels below -910HU in insp.	A
3	Air-trapping(-850). Percentage of voxels below -850HU in exp.	B
4	Ventilation. Average of ventilation values v where $v = (1000(CT_{INSP} - CT_{EXP}))/(CT_{EXP}(CT_{INSP} + 1000))$ [14]	C
5	Subtraction. Average of subtraction values s where $s = CT_{EXP} - CT_{INSP}$	C
6	Ratio. Average of ratio values r where $r = CT_{EXP}/CT_{INSP}$	C

Table 2. Calculated features. Each of the 6 listed features is calculated over 8 regions (The total lung volume (1), each lung independently (2) and each lobe independently (5)), making 48 features in total.

4.5 Classification

All classification experiments used k-Nearest-Neighbour (kNN) classifiers [3]. Briefly, each subject in the training set is plotted in feature-space based on its calculated feature values. New subjects are plotted similarly and classified based on the classes of the k nearest training subjects in feature-space.

Two types of classification were performed. Firstly 2-class classification was carried out, where subjects were classified as either having COPD (Class 1) or not (Class 0). Secondly, multi-class classification was attempted, giving each subject one of five COPD classes (0-4) as per table 1. In each case experiments were carried out using (1) a combination of all feature sets (feature sets A+B+C), (2) Only features which do not require registration (feature sets A+B) and (3) Only features which are derived from registration (feature set C). For all experiments values of k between 4 and 15 were tested and the k giving the best results was selected in each case. A leave-one-out training and testing procedure was applied in all cases.

5 Results

5.1 2-class Classification

The best results for 2-class classifications with various datasets are shown in figure 2. An ROC curve is shown in each case, with the area under the curve (Az) and the value of k used listed in the legend. The best results were obtained using all possible features (feature sets A+B+C) where an Az value of 0.92 was achieved. Using only registration-related features (feature set C) gave a reduced Az of 0.89, while using only features of the individual inspiration and expiration images (feature set C) gave the worst result with Az=0.88.

5.2 Multi-Class Classification

Multi-class classification results are illustrated in figure 3. The bar-chart illustrates for all items in each class (and overall) what percentage of the items were classified correctly and incorrectly. Incorrect classifications are further divided into 1-Class errors (meaning that the chosen COPD class was a neighbour of the true COPD class) and larger errors (mainly 2-class errors, see figure 4 for more detail). The leftmost bar in each group represents classification using the full feature set (A+B+C), the central bar using feature sets A+B and the rightmost bar using feature set C alone. Overall the best performance is achieved with feature set A+B+C, while feature set C alone achieves almost the same result. Excluding the registration-based features (using feature set A+B) gives a distinct reduction in the performance.

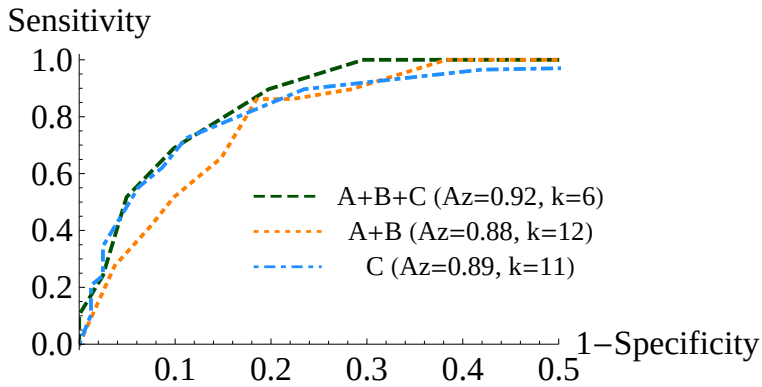


Fig. 2. ROC curves illustrating the performance of the 2-class leave-one-out classification systems with various feature sets.

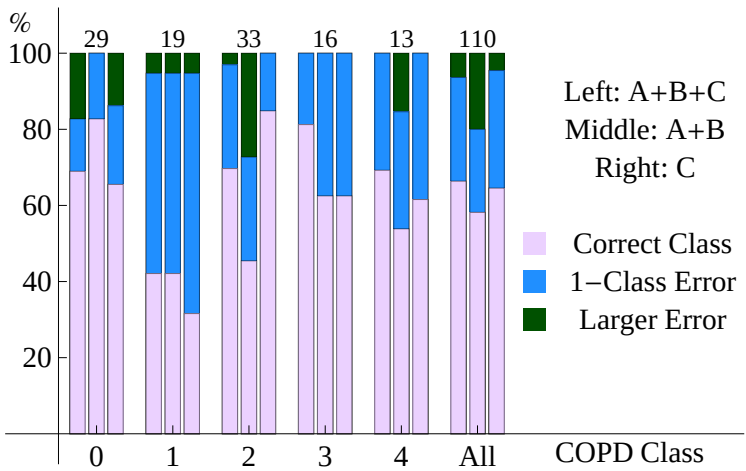


Fig. 3. Bar charts illustrating the performance of the multi-class leave-one-out classification systems with various feature sets. The number of subjects in each group is indicated by the figures above the bars.

6 Discussion

It has been shown that using automatic analysis of inspiration and expiration CT scans, subjects can be classified in a 2-class (COPD/non-COPD) system with

a high degree of accuracy (Az=0.92). Furthermore, multi-class classification is fully successful in 66% of cases, and chooses a COPD class neighbouring the correct one in a further 28%.

Precise multi-class classification of this data is difficult, partly due to the nature of the reference standard. Although pulmonary function tests are used as a gold standard to classify COPD cases they have a number of associated drawbacks and issues with reproducibility. Firstly, the test results may vary depending on how well the subject understands and adheres to the given instructions, the condition of the subject on the day of testing, the instrument used and various other factors [1]. Secondly, there is always variation in the population around the FEV1-predicted values which are based on subject height, age, weight, gender etc. [1]. Finally, since the measures used to classify COPD are on a continuous scale, even the most accurate measurements may show a subject to be very close to the border of 2 COPD classes.

In an attempt to understand the relationship between pulmonary function test scores and our system mis-classifications, all 37 cases that were misclassified in the multi-class experiment (using all features) were analysed. Figure 4 illustrates the results of this analysis with FEV1/FVC scores shown in the upper part of the figure, and FEV1/FEV1-Predicted scores in the lower part. It can be seen that the pulmonary function test scores of a number of the mis-classified cases place them very close to the boundary between two COPD classes. Cases 34, 35 and 36 (on the X-axis) for example, have FEV1/FEV1-Predicted scores placing them at the boundary between COPD classes 3 and 4. The reference standard narrowly places them in class 4, while our classifier placed them in class 3. Case number 30 has an FEV1/FEV1-Predicted score which places him in class 2, however his FEV1/FVC score is very close to the 70% boundary line, above which he would have been classified into class 0. Our classifier places this subject in class 0 indicating that his CT scans do not reveal classic signs of COPD. Further investigation would be required to determine whether the pulmonary function test scores or the CT scans of this patient truly reflect his condition. Ultimately there may be a case for using quantitative analysis of CT data in combination with results of pulmonary function tests when diagnosing COPD subjects.

All experiments illustrated that both 2-class and multi-class classification are improved by the inclusion of feature set C, based on the automatic 3D registration of the inspiration and expiration images. Direct comparison of the percentage of air present in a voxel at inspiration and at expiration is impossible without an accurate registration, and provides a qualitative measure of the pulmonary function at that precise location.

The relationship between CT-derived features and COPD classification which has been demonstrated is an extremely important one. Further research opportunities now exist to determine whether CT should be routinely used in the diagnosis and severity classification of COPD. In addition, more precise relationships between CT features and COPD can be investigated to determine, for

example, whether impairment in specific lobes affects the spirometry measurements more seriously than in others.

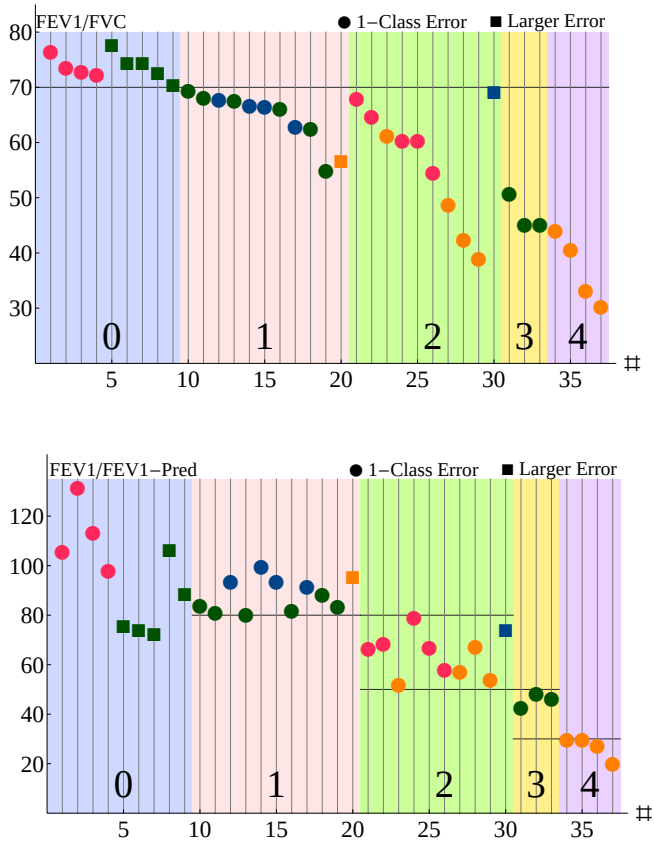


Fig. 4. Analysis of pulmonary function scores for 37 mis-classified subjects - Above: FEV1/FVC scores, Below:FEV1/FEV1-Predicted scores. Subjects are ordered by their reference standard COPD class from left to right (see background colours). The colour of each point indicates the class into which the subject was incorrectly classified. (Colours are as per the background, thus: 0=blue, 1=pink, 2=green, 3=orange, 4=purple). Horizontal lines depict the boundaries of the COPD class as defined in table 1.

References

1. M.R. Becklake. Concepts of normality applied to the measurement of lung function. *Am J Med*, 80(6):1158–1164, 1986.
2. R. Brower, R.A. Wise, C. Hassapoyannes, B. Bromberger-Barnea, and S. Permutt. Effect of lung inflation on lung blood volume and pulmonary venous flow. *J Appl Physiol*, 58(3):954–963, 1985.
3. T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Trans. Information Theory*, 13(1):21–27, 1967.
4. T. Guerrero, K. Sanders, E. Castillo, Y. Zhang, L. Bidaut, T. Pan, and R. Komaki. Dynamic ventilation imaging from four-dimensional computed tomography. *Phys Med Biol*, 51(4):777–791, 2006.
5. A. Heremans, J. A. Verschakelen, L. van Fraeyenhoven, and M. Demedts. Measurement of lung density by means of quantitative CT scanning. A study of correlations with pulmonary function tests. *Chest*, 102(3):805–811, 1992.
6. S. Hu, E.A. Hoffman, and J.M. Reinhardt. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. *IEEE Trans Med Imaging*, 20:490–498, 2001.
7. S. Klein, M. Staring, and J.P.W. Pluim. Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines. *IEEE Trans Image Process*, 16:2879–2890, 2007.
8. P. Lo, J. Sporring, H. Ashraf, J.J. Holst Pedersen, and M. de Bruijne. Vessel-guided airway segmentation based on voxel classification. In *The First International Workshop on Pulmonary Image Analysis - MICCAI*, pages 113–122, 2008.
9. Y. Nakano, S. Muro, H. Sakai, T. Hirai, K. Chin, M. Tsukino, K. Nishimura, H. Itoh, . D. Paré, J.C. Hogg, and M. Mishima. Computed tomographic measurements of airway dimensions and emphysema in smokers. *Am J Respir Crit Care Med*, 162(3):1102–1108, 2000.
10. K. F. Rabe, S. Hurd, A. Anzueto, P.J. Barnes, S.A. Buist, P. Calverley, Y. Fukuchi, C. Jenkins, R. Rodriguez-Roisin, C. van Weel, and J. Zielinski. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: Gold executive summary. *Am J Respir Crit Care Med*, 176(6):532–555, 2007.
11. J. Reilly. Using computed tomographic scanning to advance understanding of chronic obstructive pulmonary disease. *Proc Am Thorac Soc*, 3(5):450–455, 2006.
12. D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Trans Med Imaging*, 18(8):712–721, 1999.
13. A.M.R. Schilham, B. van Ginneken, H. Gietema, and M. Prokop. Local noise weighted filtering for emphysema scoring of low-dose CT images. *IEEE Trans Med Imaging*, 25:451–463, 2006.
14. B.A. Simon. Non-invasive imaging of regional lung function using X-ray computed tomography. *J Clin Monit Comput*, 16(5-6):433–442, 2000.
15. I.C. Sluimer, M. Prokop, and B. van Ginneken. Towards automated segmentation of the pathological lung in CT. *IEEE Trans Med Imaging*, 24(8):1025–1038, 2005.
16. P. Thévenaz and M. Unser. Optimization of mutual information for multiresolution image registration. *IEEE Trans. Image Proc.*, 9:2083 – 2099, 2000.
17. E.M. van Rikxoort, M. Prokop, B. de Hoop, M.A. Viergever, J.P.W. Pluim, and B. van Ginneken. Automatic segmentation of the pulmonary lobes from fissures, airways, and lung borders: evaluation of robustness against missing data. In *Medical Image Computing and Computer-Assisted Intervention (in press)*, 2009.

Automated Anatomical Likelihood Driven Extraction and Branching Detection of Aortic Arch in 3-D Chest CT

Marco Feuerstein[1], Takayuki Kitasaka[2,3], Kensaku Mori[1,3]

[1] Graduate School of Information Science, Nagoya University, Japan,
fmarco@suenaga.m.is.nagoya-u.ac.jp,
[2] Faculty of Information Science, Aichi Institute of Technology, Japan,
[3] MEXT Innovation Center for Preventive Medical Engineering, Nagoya University, Japan.

Abstract. The extraction and analysis of the aortic arch in chest computed tomography (CT) data can be an important preliminary step for the diagnosis and treatment planning of e.g. lung cancer. We here present a new method for automatic aortic arch extraction and detection of the main arterial branchings that may serve as segmentation seeds or as landmarks for intra- and interpatient registration of the mediastinum. Our method, which is based on Hough and Euclidean distance transforms and probability weighting, works on both contrast enhanced and non-contrast CT. A comparison to data manually extracted from 40 cases shows its robustness at an acceptable overall runtime.

1 Introduction

The automatic extraction of the aortic arch in computed tomography (CT) data of the chest has gained in importance in recent years, in particular to reduce the work load of physicians during diagnosis and treatment planning. As the aorta is the major vessel in the mediastinum, its automatic segmentation can be a first step for various tasks such as the definition of other mediastinal anatomy or lymph node stations for lung cancer staging. A delineation of the mediastinal vasculature for instance can be important for planning of transbronchial needle aspiration to estimate the optimal path of biopsy needles avoiding collisions with vasculature [1].

Moreover, when comparing several arches of the aorta across different data sets of the same patient or across different patients, their precise and accurate alignment is desirable to compare similarities or variances. An intuitive way to perform such a registration is to first align their arterial branchings and then refine this initial alignment towards the ascending and descending aorta.

However, the determination of the aortic arch and its branchings in CT images is not trivial, as the aortic arch frequently merges with adjacent tissue of the same image intensity, making an automation more difficult. Furthermore, only about 94.3% of all patients show a typical branching pattern, i.e. innominate, left common carotid, and left subclavian arteries, in that order [2].

The purpose of this work is two-fold. First, we present a method for fully automatic aortic arch segmentation, which works robust on both contrast enhanced and non-contrast chest CT and extends previously presented methods. Second, we utilize the segmentation result in a new algorithm to automatically determine the branchings of the aorta. These branchings can be utilized e.g. for intra- and interpatient registration of the mediastinum or as seed points for a subsequent segmentation of the branches.

Several research groups proposed methods for the extraction of the aorta in contrast enhanced data. For example, various semi-automatic and interactive methods were presented for the extraction of the abdominal aorta lumen in CT angiography data [3–5]. Behrens et al. [6] presented an approach to extract tubular structures using randomized Hough transform and Kalman filtering. It requires a starting point, a coarse direction, and an approximate radius, and was tested on the aortic arch in magnetic resonance imaging angiography data. Kovács et al. [7] utilize the Hough transform to initialize an extraction of the aortic arch in contrast enhanced CT based on a deformable surface model. O'Donnell et al. [8] and Peters et al. [9] also proposed a deformable model to fit the aorta in CT angiography data. While the former [8] was only tested on one data set, the latter [9] showed segmentation errors for lower contrast. Low contrast is a common issue for all methods developed for contrast enhanced data.

In non-contrast CT, the Hough transform was also utilized by a few groups, e.g. by Išgum et al. [10] and Kurkure et al. [11] to detect the ascending and descending part of the thoracic aorta, but not its arch or branchings. Kitasaka et al. [12] were the first to fit an aorta model to non-contrast chest CT. However, as their method only includes one model, it cannot cover a large range of variations of the aortic shape. Taeprasartsit and Higgins [13] therefore extended this approach to three models and to work on both non-contrast and contrast enhanced data. After manually selecting the carina tracheae, it could successfully extract the aorta for 12 test cases. For one case, no appropriate model could be selected.

To our knowledge, there has been no prior work on the automatic extraction of the main branchings of the aortic arch.

2 Method

Based on general a priori knowledge of the mediastinal anatomy, our method first performs a series of circular Hough transforms to delineate the aortic arch and its centerline. Using a B-spline to represent this centerline, the B-spline is then fitted to a likelihood image, which basically consists of Euclidean distances to possible aorta edges. After successful centerline fitting, the Euclidean distances along the centerline are used to recover the full segmentation of the aortic arch.

We eventually determine the main arterial branchings in a parallel projection image of the segmentation and likelihood image of the aortic arch (in superior direction) by weighing up three factors influencing the probability of branching candidates: vessel thickness (represented by the likelihood image), proximity to the projected centerline, and relative branching positions.

2.1 Preprocessing

Before any further processing, we smooth the input data by a 3-D median filter of size s_{m} (see Table 1 for all subsequent variables) to reduce image noise. Moreover, we assume the patient to be placed in true supine position, as this is the case for most chest CT acquisitions. However, if the patient is positioned differently, we could identify representative and unique anatomic landmarks such as bones [14], perform a principal component analysis or the like to estimate the main body axes, and rotate the data set accordingly.

The lung (enclosing the mediastinum) and the carina tracheae (within the mediastinum) are important anatomical landmarks, which we here utilize to estimate the first centerline points of the aorta. We automatically determine a rough segmentation of the lung area and its bounding box applying an approach similar to [15]. Within the bounding box of the lung, we search for the position of the carina by extracting the airway tree [16] and checking subsequent axial slices for the ridge between the openings of the right and left main bronchi.[4]

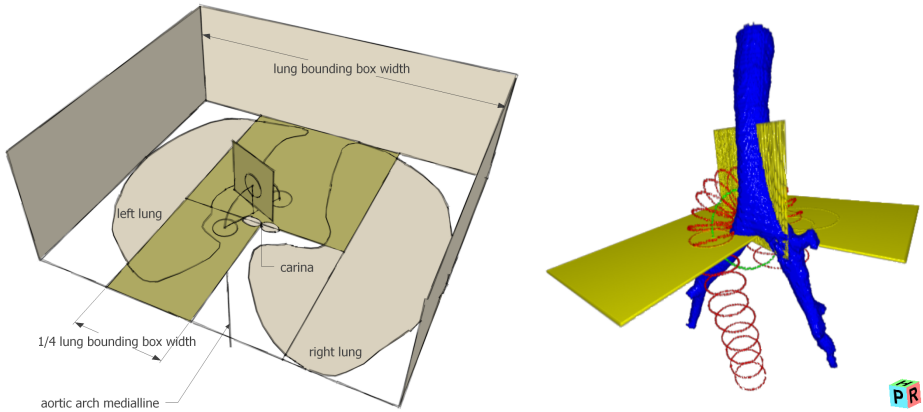
We furthermore approximate a maximum mediastinal bounding box, which is centered at the position of the carina on the mediolateral axis and at the center of the lung bounding box on the anteroposterior and superior-inferior axes. It has half the diameter (width) of the lung bounding box along the mediolateral axis and the full diameter of the lung bounding box along the anteroposterior and superior-inferior axes (cf. Fig. 1a).

2.2 Aortic Arch Segmentation

Our aortic arch segmentation method is an enhancement of the works of [7, 12, 13]. As in [12, 13], high gradients of a CT image are extracted (representing approximate blood vessel edges) and a Euclidean distance transform is applied to the gradient image to obtain the likelihood of each voxel being closer or further away form the centerline of the aorta. In [12, 13] this is followed by fitting a predetermined model of the aortic arch centerline to this likelihood image. However, to be independent of predetermined models limited to a number of cases, we automatically delineate a patient-specific aortic arch centerline by a series of Hough transforms, as done in [7], and then fit this centerline (instead of a predetermined model) to the likelihood image. The final shape of the aortic arch is then recovered by an inverse Euclidean distance transform.

Hough Circle Detection We first automatically delineate points along the centerline of the aortic arch by applying a series of circular Hough transforms. As the centerline of the aortic arch can roughly be described by a semicircle that is extended by two lines towards the inferior, we can easily restrict the Hough transformation space, once we know the radius of this semicircle. To estimate this

[4] Note that basically almost any available method for automatic lung and airway tree extraction could be chosen, as we only need a very rough approximation of the lung area and the trachea, left, and right main bronchus.



(a) Definition of the first three search regions (in khaki) for circles representing the ascending, descending, and upper part of the aortic arch.

(b) Hough circle extraction delineating the aortic arch.

Fig. 1: Hough circle extraction: (a) In two predefined search areas within the axial slice containing the carina and an oblique slice (khaki), we search for three initial Hough circles. (b) From the three centers of them another circle is computed (green). Along the upper semicircle and in the axial slices towards its inferior, we search for more Hough circles outlining the aortic arch.

radius, we search for two initial Hough circles (representing the ascending and descending aorta, respectively) inside the axial slice containing the carina and another Hough circle (representing the upper aortic arch) inside an oblique slice in between and orthogonal to the centers of the circles found in the axial slice (cf. Fig. 1a). During each Hough circle search, we restrict the search region to save computation time and avoid mis-detections. For the ascending aorta, we restrict this search region to the part of the axial slice containing the carina, which is enclosed by the mediastinal bounding box and to the anterior of the carina. For the descending aorta, we restrict it to the part enclosed by the mediastinal bounding box and to the posterior left of the carina. Once the first two circles and their centers are estimated, we can get a rough estimate of the radius of the aortic arch by computing the Euclidean distance between the two centers. We now search for a third circle in an oblique slice in between and orthogonal to the first two circle centers, which is centered the aortic arch radius away from the axial slice towards the superior. We set its size to five times the average radius of the first two circles.

During Hough circle extraction, the Hough map is computed only considering pixels between $t_{\mathrm{H}_{\mathrm{min}}}$ and $t_{\mathrm{H}_{\mathrm{max}}}$ in the input image, which is the typical edge intensity range of the aorta. For each pixel, we compute the Gaussian derivative of standard deviation σ_{H} to estimate the direction of the normal at that pixel. The Hough accumulator is filled by drawing lines between $r_{\mathrm{H}_{\mathrm{min}}}$ and $r_{\mathrm{H}_{\mathrm{max}}}$ (the typical minimum and maximum radii of the aorta) along the normal. We blur

the Hough map using a variance of v_{H} and compute the maxima in the Hough map. The highest maximum, which represents a full circle in our search area, is considered to belong to the aortic arch. The upper threshold $t_{\mathrm{H}_{\max}}$ is automatically adjusted to the average intensity of the three initial circle centers to account for contrasted data, where the range of edge voxels is much larger.

When searching for the first three Hough circles, special care needs to be taken for the circle representing the ascending aorta, as the inferior vena cava and the brachiocephalic trunk (which both can appear circular) may be visible in the same axial slice. To select the correct circle, we apply a voting, which weights the corresponding value in the Hough map, the radius of the circle (as the ascending aorta is usually the largest of the three vessels), and the distance of the circle center to the carina along the mediolateral axis (as the ascending aorta in most cases is right above the carina):

$$a = \underset{i=1\ldots n}{\arg\max} \left(\frac{h(\boldsymbol{x}_i)}{\max\limits_{i=1\ldots n}(h(\boldsymbol{x}_i))} \cdot \frac{r(\boldsymbol{x}_i)}{\max\limits_{i=1\ldots n}(r(\boldsymbol{x}_i))} \cdot \frac{d_{\mathrm{car}_{\max}} - d_{\mathrm{car}}(\boldsymbol{x}_i)}{d_{\mathrm{car}_{\max}}} \right) \tag{1}$$

where $h(\boldsymbol{x}_i)$ is the value in the Hough map corresponding to circle i, $r(\boldsymbol{x}_i)$ is the radius of this circle, $d_{\mathrm{car}}(\boldsymbol{x}_i)$ is the distance of the circle center to the carina in mediolateral direction, and $d_{\mathrm{car}_{\max}}$ is half the mediastinal bounding box diameter in mediolateral direction. For the descending aorta, we just take the circle corresponding to the maximum value in the Hough map. For the upper aortic arch circle, we need to take care of the left pulmonary artery, which sometimes runs parallel below the upper aortic arch. Hence we perform another voting for the most probable circle:

$$u = \underset{i=1\ldots n}{\arg\max} \left(\frac{h(\boldsymbol{x}_i)}{\max\limits_{i=1\ldots n}(h(\boldsymbol{x}_i))} \cdot \frac{r(\boldsymbol{x}_i)}{\max\limits_{i=1\ldots n}(r(\boldsymbol{x}_i))} \cdot \frac{d_{\mathrm{cen}_{\max}} - d_{\mathrm{cen}}(\boldsymbol{x}_i)}{d_{\mathrm{cen}_{\max}}} \right) \tag{2}$$

where $d_{\mathrm{cen}}(\boldsymbol{x}_i)$ is the Euclidean distance of the circle center to the slice center and $d_{\mathrm{cen}_{\max}}$ is half the length of the diagonal of the oblique slice.

From the three center points of the Hough circles, we estimate a circle in 3D, its upper semicircle representing an estimate for the upper part of the aortic arch (cf. Fig. 1b).

Following the approach of [7], we reconstruct oblique 2-D slices of four times the size of the average radius of the three initial Hough circles in 15° steps along the semicircle. In each oblique slice, we search for the Hough circle with the maximum value in the corresponding Hough map. At each of the two ends of the semicircle, we start a search for the ascending and descending parts of the aortic arch. Every 12.5 mm we determine Hough circles in axial slices of the same size as before, iteratively adjusting the center of the current slice to the center coordinates of the previous Hough circle and adjusting the size of the 2-D slices to be four times the size of the average radius of the last three extracted circles (we hence take care for the facts that the aorta is not straight and the radius in the descending part is decreasing). We perform this process for the ascending aorta twice and for the descending aorta eight times. Overall, depending on the size of the data set, we get at most 23 initial centerline points.

NURBS Fitting From the initial centerline points, we generate a nonuniform rational B-spline (NURBS) curve that best fits the points in a least squares sense. We use 15 control points for the NURBS curve, so even if some of the centerline points are a little off, we get a good first approximation.

To match the curve with the real centerline of the aortic arch, we create a likelihood image of the centerline according to [12] (cf. Fig. 2a). In detail, we first perform a morphological opening using a sphere of size r_{A} on the median-filtered image to reduce variance of voxel intensities. Next, we detect edges in the opened image by computing the gradient magnitude and only leave voxels with a magnitude greater than $t_{\Delta_{\mathrm{E}}}$. As the variance of intensities inside blood vessels is usually low, we can reduce false edge candidates by computing the standard deviation at each edge candidate within a sphere of radius $r_{\sigma_{\mathrm{E}}}$ in the opened image and only leaving candidates whose standard deviation is greater than $t_{\sigma_{\mathrm{E}}}$. Last, to generate the final likelihood image we apply a Euclidean distance transform to the edge image supplemented with "artificial" edge voxels obtained from all voxels of the rough lung segmentation. The likelihood image shows the distance between a voxel and its nearest edge voxel and thus the likelihood of a voxel to be part of the centerline. Using additional "artificial" edge voxels we ensure that all air voxels have zero likelihood to be part of the aorta.

In contrast enhanced data, the intensity distribution varies a lot within the contrasted region, leading to unwanted edges inside the aorta. To only consider its wall, we compute the average image intensity at the initial centerline points. If the average intensity exceeds a threshold t_{C}, we adjust all input voxels to be smaller or equal to t_{C} before performing above likelihood image generation steps.

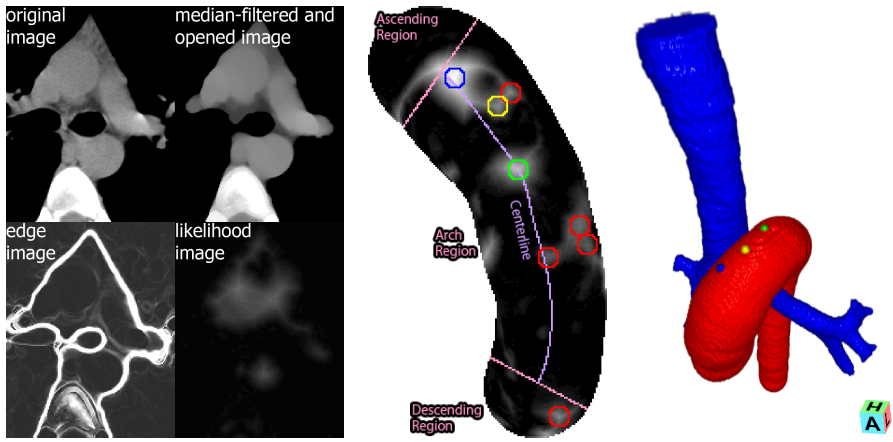
Next, using the Powell optimizer, the NURBS curve is fitted to the likelihood image by minimizing following expression:

$$\arg\min_{\boldsymbol{P}_i} \left(-\frac{1}{m} \sum_{j=1}^{m} d_{\mathrm{L}}^2 \left(N\left(\frac{j}{m}\right)\right)\right), \text{ where } N(u) = \sum_{i=1}^{k} R_{i,p}\boldsymbol{P}_i \qquad (3)$$

Here, $d_{\mathrm{L}}(\boldsymbol{X})$ is the Euclidean distance value of the voxel $\boldsymbol{X}$ in the likelihood image, N is the NURBS curve, m is the number of sampling points along the curve (we sample every millimeter), $R_{i,p}$ are the rational basis functions of the curve (of degree $p = 3$ in our case), and $\boldsymbol{P}_i = (x_i, y_i, z_i)^T$ is the ith of k control points. Compared to [12] and [13], where one or more models need to be fitted globally as well as locally using several energy terms, we here greatly reduce the amount and complexity of optimizations to a single local minimization and energy term, making the approach faster and stable.

Finally, we recover the shape of the aortic arch by a reverse Euclidean distance transform, followed by a procedure to deal with false edges. In detail, we initialize each centerline voxel with the corresponding Euclidean distance from the likelihood image and draw a sphere of radius equal to this distance. As the likelihood image may still contain wrong edges, we grow each sphere iteratively, until the standard deviation of all voxels within the sphere exceeds $t_{\sigma_{\mathrm{R}}}$. Each

iteration increases the radius of the sphere by the smallest of the three spatial resolutions of the input image.



(a) Likelihood image generation steps.

(b) Branching assignment (inferior view).

(c) Extracted branchings.

Fig. 2: Processing steps during branching assignment: (a) Generation of a 3-D likelihood image. (b) 2-D projection of the likelihood image and the centerline of the aortic arch segmentation (purple), and division of the projection into ascending, arch, and descending region by the pink lines. Each branching candidate is a local intensity maximum within a certain radius r_B represented by the colored circle. Red are unassigned candidates, blue are candidates assigned to the innominate artery, yellow to the left common carotid artery, and green to the left subclavian artery. (c) Final branching assignment results.

2.3 Branching Extraction

The extraction of all branchings is motivated by the fact that the innominate, left common carotid, and left subclavian arteries branch off the aortic arch in superior direction. We incorporate three properties of the branchings into our algorithm. First, we search for voxels along the boundary of the aortic arch, which have a higher local Euclidean distance (likelihood) than others. Second, the closer the voxels are to the upper ridge of the aortic arch, the more likely they are branchings. And third, the branching arteries usually have a certain distance to each other.

Preprocessing Before starting our branching extraction, we preprocess the input image in the same way as for aortic arch segmentation to obtain a likelihood image (cf. Fig. 2a), including thresholding with t_C in case of contrast enhanced data. However, as the average diameter of the aortic arch branches is smaller

than that of the arch itself, we reduce the radius of the structuring element for morphological opening to r_{B} in order not to "smooth away" small branches.

Parallel Projection After preprocessing, we create the (one-voxel thick) boundary of the upper aortic arch segmentation. To simplify our branching search, we reduce the search space from three to two dimensions: starting from the axial slice containing the carina, we perform a parallel projection of the likelihood voxels inside the boundary onto a single image, in the following referred to as 2-D likelihood image. Furthermore, for the 2-D likelihood image, we compute the 2-D Euclidean distance transform to its boundary. This image, in the following referred to as 2-D boundary distance image, is an indicator for the likelihood of a candidate to be a branching, as the upper ridge of the aortic arch roughly corresponds to high values in the 2-D boundary distance image.

At the same time we perform a parallel projection of the segmented centerline of the aortic arch and approximate a 2-D NURBS curve n to this centerline projection, which again roughly overlaps with the upper ridge of the arch. For each point $\boldsymbol{x}$ in the 2-D likelihood image or the 2-D boundary distance image, we define its offset $\delta(\boldsymbol{x})$ to the first control point of n along the curve by

$$\delta(\boldsymbol{x}) = \begin{cases} -|n(0) - \boldsymbol{x}| & \text{if } f(\boldsymbol{x}) = 0 \quad \textbf{(ascending region)} \\ l_{\mathrm{n}}(0, f(\boldsymbol{x})) & \text{if } 0 < f(\boldsymbol{x}) < 1 \quad \textbf{(arch region)} \\ l_{\mathrm{n}}(0,1) + |n(1) - \boldsymbol{x}| & \text{if } f(\boldsymbol{x}) = 1 \quad \textbf{(descending region)} \end{cases} \tag{4}$$

where $l_{\mathrm{n}}(a, b)$ gives the arc length along n between a and b and $f(\boldsymbol{x})$ gives the parameter of the point on the curve closest to $\boldsymbol{x}$. Using $\delta(\boldsymbol{x})$, we can also assign one of three approximate regions (ascending, arch, descending) to each point (cf. Fig. 2b).

Branching Assignment In the 2-D likelihood image, we search for local maxima greater than r_{B} within a radius of r_{B}, whose neighbors within r_{B} are all inside the boundary of the 2-D likelihood image. These maxima represent our initial branching candidates, which we now need to identify the correct ones from and assign correct arteries to.

Innominate Artery As the innominate artery branching may be located at many possible positions along the upper aortic arch and the innominate and left subclavian artery may have very similar radii and thus likelihoods, we cannot simply select the innominate as the candidate with the highest likelihood. We first need to restrict our search area along the upper aortic arch. Therefore, we compute the average weighted distance d_{w} of all candidates to the first control point of n by

$$d_{\mathrm{w}} = \frac{\sum_{j=1}^{w} \delta(\boldsymbol{x}_j) \cdot d_{\mathrm{l}}(\boldsymbol{x}_j)}{\sum_{j=1}^{w} d_{\mathrm{l}}(\boldsymbol{x}_j)} \tag{5}$$

where $d_{\mathrm{l}}(\boldsymbol{x})$ is the Euclidean distance value of the pixel $\boldsymbol{x}$ in the likelihood image. The candidate positioned below d_{w} (i.e. towards the ascending region) with the highest likelihood is assigned to the innominate artery.

However, the innominate artery is often adjacent to the left innominate vein, so sometimes no distinct border between them is visible, which can lead to several local likelihood maxima. Therefore, if it is located in the ascending region along with other candidates, we update the index i of the most probable candidate to

$$i = \arg\max_{j=1\ldots w} \left(\frac{d_{\mathrm{l}}(\boldsymbol{x}_j)}{\max\limits_{j=1\ldots w}(d_{\mathrm{l}}(\boldsymbol{x}_j))} \cdot \frac{d_{\mathrm{b}}(\boldsymbol{x}_j)}{d_{\mathrm{b}}(n(0))} \right) \tag{6}$$

where $d_{\mathrm{b}}(\boldsymbol{x})$ is the value of the pixel $\boldsymbol{x}$ in the boundary distance image and w is the number of candidates. This favors posterior candidates with higher likelihood.

Left Subclavian Artery As we got some of the branching candidates due to merges of the aortic arch with adjacent tissue, and the thickness of the left subclavian and left common carotid arteries can be similar and much smaller than that of the innominate artery, we cannot simply search for the second largest value in the 2-D likelihood image. We need to consider their natural positions relative to each other.

So first we compute the offset $d_{\mathrm{i}}(\boldsymbol{x}_j) = \delta(\boldsymbol{x}_j) - \delta(\boldsymbol{x}_i)$ between the remaining candidates and the branching of the innominate artery. If the remaining candidates lie posterior to the innominate artery branching and are at most the arc length $l_{\mathrm{n}}(0,1)$ of the whole centerline curve away, we further consider them in the following expression to get the index s of the most likely candidate for the left subclavian artery:

$$s = \arg\max_{j=1\ldots v} \left(\frac{d_{\mathrm{l}}(\boldsymbol{x}_j)}{\max\limits_{j=1\ldots v}(d_{\mathrm{l}}(\boldsymbol{x}_j))} \cdot \frac{d_{\mathrm{b}}(\boldsymbol{x}_j)}{d_{\mathrm{b}}(n(f(\boldsymbol{x}_j)))} \cdot \left(1 - \frac{|l_{\mathrm{n}}(0,\frac{1}{3}) - d_{\mathrm{i}}(\boldsymbol{x}_j)|}{l_{\mathrm{n}}(0,\frac{2}{3})} \right) \right) \tag{7}$$

Here we account for the facts that the likelihood should be as high as possible, the candidate should be as close to the centerline as possible, and the branching of the left subclavian artery should be about one third the arc length of the centerline curve away from the branching of the innominate artery.

Left Common Carotid Artery All u remaining candidates are considered in a final step, if they lie in between the innominate and left subclavian artery (i.e. $0 < d_{\mathrm{i}}(\boldsymbol{x}_j)$ and $0 < d_{\mathrm{s}}(\boldsymbol{x}_j) = \delta(\boldsymbol{x}_s) - \delta(\boldsymbol{x}_j)$), are approximately on a line connecting the innominate and left subclavian artery (i.e. the angle between this line and the line connecting the candidate and the left subclavian artery should not get too big, in our case less than 40 degrees), and are superior to the inferior of the two (looking at their original depth in 3D). From the indexes of these candidates, we choose the most likely one c for the left common carotid artery by

$$c = \arg\max_{j=1\ldots u} \left(\frac{d_{\mathrm{l}}(\boldsymbol{x}_j)}{\max\limits_{j=1\ldots u}(d_{\mathrm{l}}(\boldsymbol{x}_j))} \cdot \frac{d_{\mathrm{b}}(\boldsymbol{x}_j)}{d_{\mathrm{b}}(n(f(\boldsymbol{x}_j)))} \cdot \left(1 - \frac{|d_{\mathrm{s}}(\boldsymbol{x}_j) - d_{\mathrm{i}}(\boldsymbol{x}_j)|}{d_{\mathrm{s}}(\boldsymbol{x}_j) + d_{\mathrm{i}}(\boldsymbol{x}_j)} \right) \right) \tag{8}$$

Here we favor candidates, which lie half way between the branchings of the innominate and the left subclavian artery.

3 Results

We evaluated our method on 10 contrast enhanced and 30 non-contrast chest CT data sets of various hospitals, scanners, and acquisition parameters. The 40 data sets consisted of 99 – 838 slices spaced 0.4 – 1.5 mm. Each axial slice had 512 × 512 pixels of size 0.5 – 0.665 mm. For each data set, we manually segmented the aortic arch and extracted its branchings, which took approximately one to two hours per data set. For the automatic algorithm, we used the parameters shown in Table 1.

Table 1: Values of parameters used for evaluation (HU refers to Hounsfield units).

s_m: 3×3×3 voxels		$t_{H_{min}}$: -60 HU		v_H: 5 pixels	t_{Δ_E}: 10 HU
t_C:	200 HU	$t_{H_{max}}$: 40 HU		σ_H: 5 pixels	r_{σ_E}: 3 voxels
r_A:	4 mm	$r_{H_{min}}$: 7 mm			t_{σ_E}: 15 HU
r_B:	2.5 mm	$r_{H_{max}}$: 28 mm			t_{σ_R}: 8 HU

For each data set, we computed the sensitivity, specificity, and Jaccard index between the manually and automatically segmented 3-D volumes (considering voxels within the bounding box of the two segmentations only) and their mean, standard deviation (SD), and minimum. We also computed the mean, SD, and maximum of the minimum Euclidean distances (MD) of the contours of the two segmentations and their mean, SD, and maximum. Furthermore, we counted the number of false positives (i.e. branchings outside their corresponding artery) and false negatives (i.e. missed branchings) for all 40 data sets and computed the mean, standard deviation, and maximum of the Euclidean distances between all manually and automatically selected branchings. Finally, on a workstation with two 64-bit Quad-Core Intel Xeon 5355 processors and 16 GB main memory we measured the runtime of our automatic method, separated into preprocessing, aorta, branching extraction, and total runtime, along with their mean, SD, and maximum. Table 2 summarizes all results.

Table 2: Results of our evaluation.

Aortic Arch Segmentation

Sensitivity	Specificity	Jaccard Index	Mean MD (mm)	SD MD (mm)	Max MD (mm)
0.95±0.03≥0.89	0.99±0.00≥0.98	0.92±0.02≥0.85	0.4±0.1≤0.9	0.5±0.1≤0.9	3.6±1.2≤6.9

Branching Extraction			Runtime (s)			
Distance (mm)	TP	FP	FN	Preprocessing	Aortic Arch	Branching	Total
2.0±1.1≤6.1	114	0	3	68±23≤106	74±42≤179	12±5≤21	154±65≤298

4 Discussion

Our results show that, at an acceptable average total runtime of about 2.5 minutes and a mean distance of less than half a millimeter between manual and automatic segmentation, the aortic arch could be extracted well. As can be seen in Table 3, we also improve the state of the art in aortic arch segmentation.

As already mentioned in [12, 13], problems can arise when cardiac motion or calcifications induce imaging artifacts and when the pulmonary artery, superior

Table 3: Comparison of our method to the state of the art in automatic aortic arch extraction. In [13], results were not assessed quantitatively. Due to global and local fitting to one or several models, [12] and [13] are expected to have a longer runtime than our method.

Method	Non-/Contrast	Successful/Cases	Mean MD (mm)	Runtime (min)
Kovács et al. [7]	No/Yes	17/21	1.1±0.2	∼6
Peters et al. [9]	No/Yes	36/37	0.6±1.1	0.2
Kitasaka et al. [12]	Yes/No	7/7	0.5±0.2	NA
Taeprasartsit & Higgins [13]	Yes/Yes	12/13	NA	NA
Ours	Yes/Yes	40/40	0.4±0.1	∼2.5

vena cava, or other tissue is adjacent to the aorta, leading to slight overlaps and mis-extractions. However, our quantitative analysis (cf. Table 2) shows that all mis-extractions are minor and do not influence the performance of our branching extraction significantly.

The mean distance between manually and automatically extracted branchings was 2 millimeters. Only 3 out of 117 branchings could not be found by our algorithm. Similar to aortic arch extraction, this usually happened when the left common carotid artery was too close to one of the others and in the presence of calcifications or imaging artifacts, so no distinct local likelihood maximum could be found.

In 3 of our 40 cases, the common carotid artery was branching off the innominate artery instead of the aortic arch. Our method handled all these cases correctly by only assigning two branchings. In about 4.6% of a larger study [2], four arteries branched off the aortic arch. As none of our 40 evaluation cases covered such a branching pattern, we need to further analyze the common location of the fourth artery and integrate it into our method, e.g. by utilizing pattern classification techniques.

5 Conclusion

We developed a new method for the automatic detection of the main arterial branchings of the aortic arch, based on a robust technique for automatic aortic arch extraction in chest CT that extends and improves the current state of the art [7, 12, 13]. Our method works stable on both contrast enhanced and non-contrast CT, making it applicable to a large number of data sets. It can support the physician's diagnosis and treatment planning and provides valuable landmarks for further segmentation of the aortic branches, intra- and interpatient registration of the mediastinum, or chest atlas generation.

Acknowledgments

Parts of this research were supported by the Japan Society for the Promotion of Science (JSPS) postdoctoral fellowship program for foreign researchers, a Grant-In-Aid for Scientific Research from JSPS, the program of formation of innovation center for fusion of advanced technologies "Establishment of early

preventing medical treatment based on medical-engineering for analysis and diagnosis" funded by the Ministry of Education (MEXT), and a Grant-In-Aid for Cancer Research from the Ministry of Health, Labour and Welfare.

References

1. Taeprasartsit, P., Higgins, W.E.: System for definition of the central-chest vasculature. In: SPIE Medical Imaging. (2009)
2. Nelson, M.L., Sparks, C.D.: Unusual aortic arch variation: Distal origin of common carotid arteries. Clinical Anatomy **14** (2001) 62–65
3. Subramanyan, K., Steinmiller, M., Sifri, D., Boll, D.: Automatic aortic vessel tree extraction and thrombus detection in multislice CT. In: SPIE Medical Imaging. (2003)
4. de Bruijne, M., van Ginneken, B., Viergever, M.A., Niessen, W.J.: Interactive segmentation of abdominal aortic aneurysms in CTA images. Medical Image Analysis **8**(2) (2003) 127–138
5. Olabarriaga, S.D., Rouet, J.M., Fradkin, M., Breeuwer, M., Niessen, W.J.: Segmentation of thrombus in abdominal aortic aneurysms from CTA with nonparametric statistical grey level appearance modeling. IEEE Transactions on Medical Imaging **24**(4) (2005) 477–485
6. Behrens, T., Rohr, K., Stiehl, H.S.: Robust segmentation of tubular structures in 3-D medical images by parametric object detection and tracking. IEEE Trans Syst Man Cybern B Cybern. **33**(4) (2003) 554–561
7. Kovács, T., Cattin, P., Alkadhi, H., Wildermuth, S., Székely, G.: Automatic segmentation of the vessel lumen from 3D CTA images of aortic dissection. In: Bildverarbeitung für die Medizin. (2006)
8. O'Donnell, T., Gupta, A., Boult, T.: A new model for the recovery of cylindrical structures from medical image data. In: CVRMed-MRCAS. (1997)
9. Peters, J., Ecabert, O., Lorenz, C., von Berg, J., Walker, M.J., Ivanc, T.B., Vembar, M., Olszewski, M.E., Weese, J.: Segmentation of the heart and major vascular structures in cardiovascular CT images. In: SPIE Medical Imaging. (2008)
10. Išgum, I., Rutten, A., Prokop, M., van Ginneken, B.: Detection of coronary calcifications from computed tomography scans for automated risk assessment of coronary artery disease. Medical Physics **34**(4) (2007) 1450–1461
11. Kurkure, U., Avila-Montes, O., Kakadiaris, I.: Automated segmentation of thoracic aorta in non-contrast CT images. In: ISBI. (2008)
12. Kitasaka, T., Mori, K., Hasegawa, J., Toriwaki, J., Katada, K.: Automated extraction of aorta and pulmonary artery in mediastinum from 3D chest X-ray CT images without contrast medium. In: SPIE Medical Imaging. (2002)
13. Taeprasartsit, P., Higgins, W.E.: Method for extracting the aorta from 3D CT images. In: SPIE Medical Imaging. (2007)
14. Kunz, P., Scholz, M., Haas, B., Coradi, T.: Anatomic orientation in medical images (2008)
15. Hu, S., Hoffman, E.A., Reinhardt, J.M.: Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. IEEE Transactions on Medical Imaging **20**(6) (2001) 490–498
16. Kitasaka, T., Mori, K., Suenaga, Y., Hasegawa, J., Toriwaki, J.: A method for segmenting bronchial trees from 3D chest X-ray CT images. In: MICCAI. (2003)

Segmentation-Based Quantitation of Pulmonary Alveolar Proteinosis, Pre- and Post-Lavage, Using High-Resolution Computed Tomography

Tessa Sundaram Cook[1], Nicholas Tustison[1], Gang Song[2], Suyash Awate[1], Drew A. Torigian[1], Warren Gefter[1], and James C. Gee[1]

[1] Dept. of Radiology, Hospital of the University of PA, USA
[2] Dept. of Computer Science, University of Pennsylvania, USA

Abstract. Pulmonary alveolar proteinosis is a disorder of surfactant deposition that leads to the accumulation of lipoproteinaceous material in the alveoli. The mainstay of treatment is whole-lung lavage, by which saline is used to wash the alveolar deposits from the lung. Patients are currently followed using sequential pre- and post-lavage CT to subjectively assess treatment efficacy. We present an approach to quantify global and regional changes in disease distribution before and after lavage using automated and semiautomated segmentation methods. Histogram analysis demonstrates a leftward shift in parenchymal intensities after lavage, reflecting an increase in the degree of aerated lung after removal of excess surfactant. This methodology can be applied to a variety of pulmonary pathologies that manifest as a combination of alveolar and interstitial disease, to enable more quantitative assessment of disease progression and treatment efficacy.

1 Introduction

Pulmonary alveolar proteinosis (PAP) is a rare but potentially devastating condition in which lipoproteinaceous deposits accumulate in the alveoli as a result of impaired surfactant clearance, [1]. Congenital, idiopathic and acquired forms exist, and recent studies have linked PAP to a relative deficiency of granulocyte-macrophage colony stimulating factor (GM-CSF), a substance thought to activate alveolar macrophages [2]. However, pharmacologic treatment for PAP remains an active area of research. Currently, the mainstay of treatment is whole-lung lavage (WLL), which is performed under general anesthesia and involves sequentially washing out one lung with many liters of normal saline while ventilating the other, [3]. Multiple treatments are often required over a patient's lifetime, as alveolar deposits build up between treatments, and pneumonia with atypical organisms can cause death in these patients.

Computed tomography (CT) has been used to qualitatively assess the success of WLL. PAP has traditionally been associated with the "crazy paving" pattern of thickened interlobular septa and ground glass opacification, [4], although findings span the spectrum from alveolar predominance to interstitial fibrosis, [5].

Evaluation of CT studies by radiologists typically involves visually estimating a global increase or decrease in the volume of diseased lung. Reports of these studies generally assess the extent of disease as "same", "worse" or "better".

We present an approach toward quantifying the effects of whole-lung lavage in the treatment of pulmonary alveolar proteinosis. Using a combination of automated and semiautomated segmentation, we estimate the volume of lung parenchyma affected by disease and quantify both global and lobar effects of lung lavage on five patients with PAP. Our approach is an initial step toward an objective assessment of the efficacy of WLL—an improvement upon subjective interpretations currently used in clinical practice by radiologists.

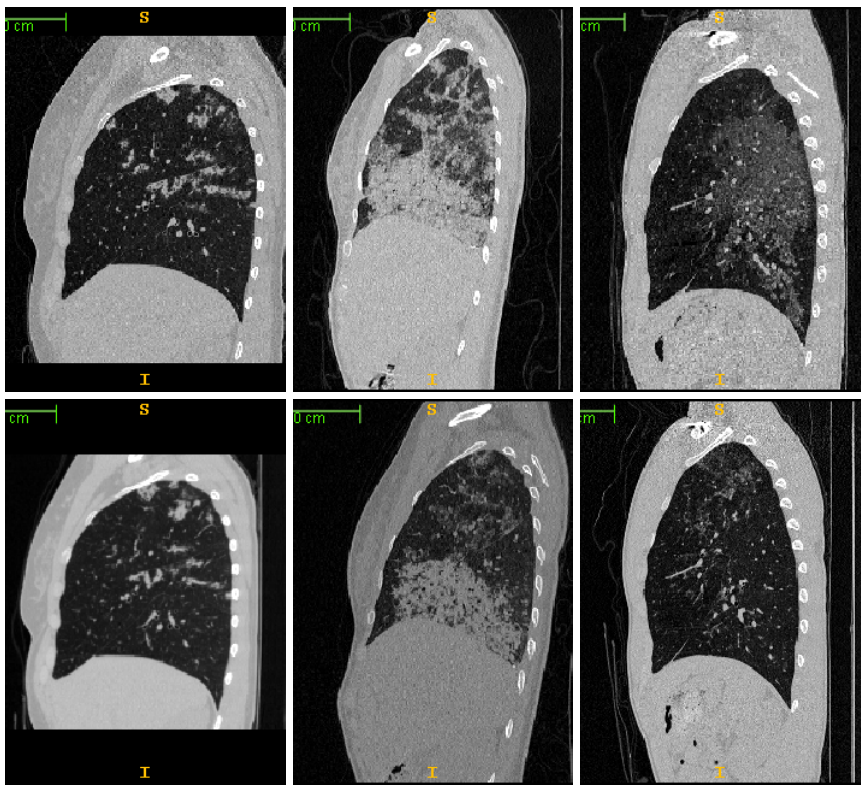


Fig. 1. Sagittal sections pre- (top) and post-lavage (bottom) show examples of lung involvement in three individuals, patients A, C and D. In some patients (e.g., A and D), lavage can be extremely effective, while in others (e.g., C) there is not much difference between the pre- and post-treatment images.

2 Materials and Methods

The complete analysis algorithm is illustrated in figure 2, using representative coronal CT sections from patient A both before and after treatment with whole-lung lavage. In the sections that follow, the terms "intensity" and "CT attenuation" are used interchangeably.

2.1 Data

High-resolution CT (HRCT) scans from five patients (designated A–E) with pulmonary alveolar proteinosis obtained immediately prior to and several months after WLL were selected for analysis. The pre- to post-lavage imaging interval ranged from 2 weeks to 12 months. The data were acquired with a matrix of 512×512 and sub-millimeter slice thicknesses. To stay within the memory limitations of our image registration algorithm (applied in the next section), images are resampled to dimensions of $256 \times 256 \times N$, where N is the number of slices required to achieve isotropic voxels. Figure 1 shows sagittal sections pre- and post-lavage in three patients used in this experiment.

2.2 Whole-Lung and Lobar Segmentation

The pre-lavage images are segmented using an automated segmentation pipeline, implemented in the open-source Insight Toolkit, [6], that employs the methodology of Hu et al, [7]. First, an optimal threshold is calculated to separate the airways and lungs from the rest of the body. Whereas in [7] an iterative approach is used, we achieve better results more quickly using Otsu thresholding. Following the segmentation of the lungs and airways from the body, we isolate the trachea from a proximal axial slice using a Hough transform of the region of interest. We then iterate through subsequent slices to segment the remainder of the trachea by propagating the solution at the previous slice to the current slice. This iterative process stops once we have propagated the solution into both the left and right lungs, yielding the segmentation of three anatomic regions of interest (trachea, left lung, right lung). Finally, smoothing of the segmentation is performed using a specific ordering of binary morphological operations as suggested in [7].

Lobar segmentation follows in a semi-automated fashion using ITK-SNAP,[8]. The active contours are initialized using edge-based parameters rather than intensity-based thresholds, and segmentation is advanced in a stepwise fashion to generate a gross approximation of a particular lung lobe. Manual editing using the knife and polygon tools is then used to complete the segmentation.

The post-lavage images are segmented by registration of the pre- and post-lavage images and warping of the pre-lavage segmentation into the domain of the post-lavage image. For registration, we use the open-source Advanced Normalization Tools (ANTS), [9], which offer several similarity metrics and both linear and non-linear transformation options for accurate modeling of biomechanical deformations. For the experiments discussed in this paper we use the

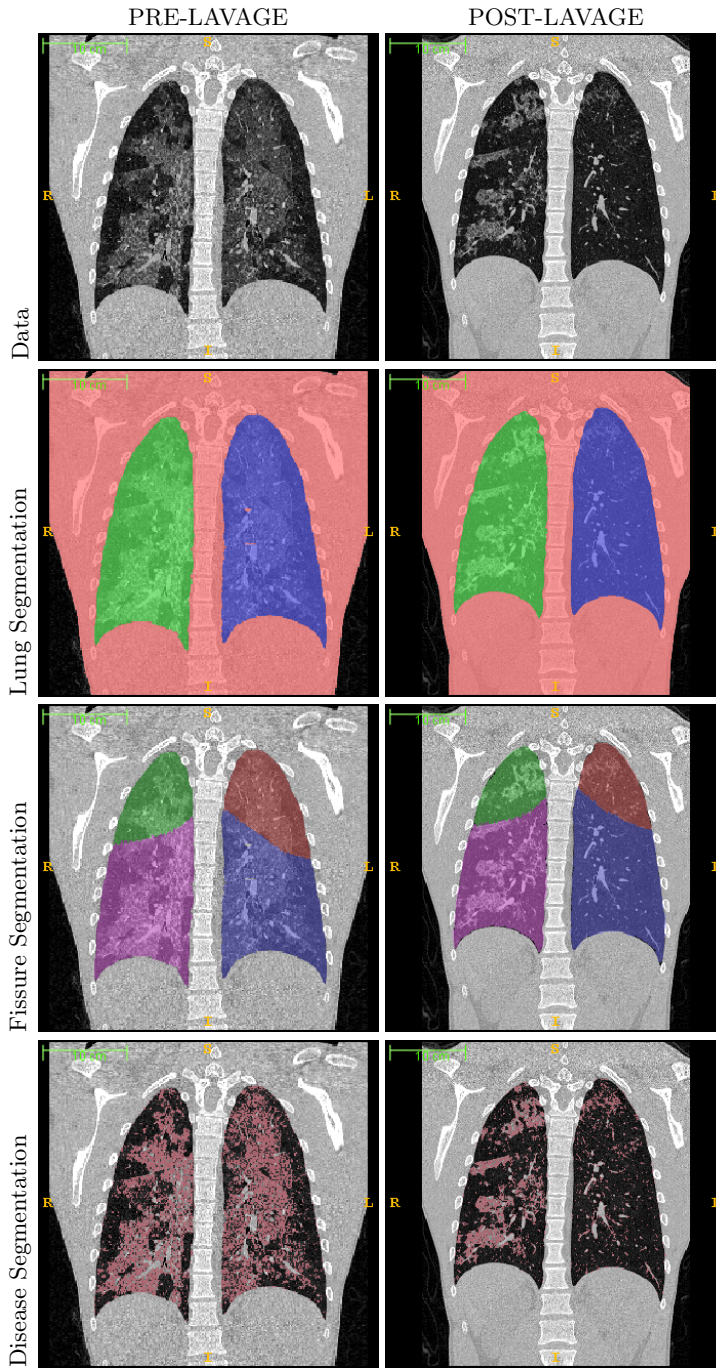


Fig. 2. The complete analysis algorithm illustrated using representative coronal sections from a PAP patient, pre- (left) and post-lavage (right). Details of each step can be found in section 2.

cross correlation similarity metric and Gaussian-regularized symmetric normalization (SyN) transformation model described in [10], which yields both the forward and inverse deformation fields after affine initialization.

2.3 Disease Segmentation Using Severity-Based Thresholds

Segmentation of regions of diseased lung is performed by empirically choosing thresholds based on the severity of the patient's disease. Patients with less severe or less concentrated PAP intrinsically ventilate a higher percentage of their alveoli, and require thresholds closer to -1000 Hounsfield units (HU), the attenuation of air, to segment alveolar deposits. Conversely, patients with extensive disease or very concentrated disease require thresholds closer to -100 HU, an attenuation value between fat and water, since very little air mixes with the surfactant accumulating in their alveoli. For each of the five patients, thresholds were customized to the severity of their disease, as determined by visual inspection of the degree of alveolar infiltration. The same intensity thresholds were used for both the pre- and post-lavage images.

2.4 Analysis

Whole-lung and lobar volumes were computed from each patient's image segmentations before and after WLL. Subsequently, the volume of disease in each lung was computed after the disease-specific thresholds were applied. The volumetric percentage of PAP in the lungs as a whole as well as in each lobe was then calculated, and pre- and post-lavage percentages were compared. Histogram analysis of the distribution of CT attenuation values in the lungs before and after treatment–normalized to the instantaneous lung volume during breath-holding–was also performed, with the expectation that after lavage, patients would demonstrate improved lung aeration, or an increased number of voxels closer to -1000 HU.

3 Results

Patients A and B were classified as having moderate disease, because their alveolar deposits, though distributed throughout all lobes, were not extremely dense. Intensities in $[-750, -300]$ HU were used to segment disease in these two patients. Patient C was noted to have focal, dense surfactant accumulation which, though it did not involve large portions of lung, was found to be extremely concentrated. Intensity thresholding in $[-200, 0]$ HU was used for patient C, reflecting the high concentration of lipid (normally around -120 HU). Patients D and E were classified as having severe disease, because their alveolar deposits were not only widespread in both lungs but also fairly dense. In these two patients, intensities in $[-600, 0]$ HU best captured regions of disease.

Results of the pre- and post-lavage analysis are summarized in table 1. For each lobe of the lung and both lungs as a whole, we report the volume of disease

Region	Time	A	B	C	D	E
LUL	*pre*	0.31	0.63	0.10	0.16	0.49
	post	0.18	0.39	0.11	0.09	0.31
LLL	*pre*	0.47	0.39	0.09	0.24	0.64
	post	0.14	0.23	0.08	0.14	0.39
RUL	*pre*	0.23	0.41	0.13	0.15	0.18
	post	0.22	0.37	0.13	0.08	0.17
RML	*pre*	0.13	0.47	0.06	0.37	0.26
	post	0.11	0.29	0.07	0.21	0.18
RLL	*pre*	0.44	0.44	0.09	0.34	0.64
	post	0.27	0.27	0.09	0.19	0.47
Whole	*pre*	0.35	0.46	0.10	0.24	0.47
	post	0.19	0.30	0.10	0.13	0.33

Table 1. Summary of the global and lobar disease percentages (volume of disease in the region divided by total region volume) before and after WLL in patients A-E. LUL=left upper lobe, LLL=left lower lobe, RUL=right upper lobe, RML=right middle lobe, RLL=right lower lobe, Whole=both lungs.

in the region divided by the volume of the region itself before and after treatment. With the exception of patient C, who demonstrated nearly identical volumes of disease pre- and post-lavage, we quantify at least a 10% reduction in whole-lung disease in each patient, with decreases of up to 70% noted within individual lobes. As an example, 3-D renderings of disease distribution pre- and post-treatment for patient D (severe disease), are shown in figure 3.

Histograms of the CT attenuation in the lungs normalized to lung volume show that most patients who demonstrated a qualitative improvement in disease extent on CT experienced a leftward shift in parenchymal intensity distribution after lavage. This suggests a combination of factors: the expected improvement in lung aeration after removal of alveolar deposits as well as the ability to maintain a deeper inspiration post-treatment. Figures 4 and 5 demonstrate a marked leftward shift post-WLL in patient D, who had severe disease, both on a lobar level and over both lungs as a whole. These findings can be visually correlated with the 3-D renderings in figure 3. Comparatively, patient C, the patient with the dense though sparsely distributed areas of disease, did not appreciably respond to lavage, reflected both in the quantitative analysis (table 1) and in figure 6.

4 Discussion

The current clinical standard for assessment of disease in PAP is subjective analysis of increased or decreased alveolar involvement using serial CT over a patient's lifetime. In this work, we explore a quantitative approach to objectively

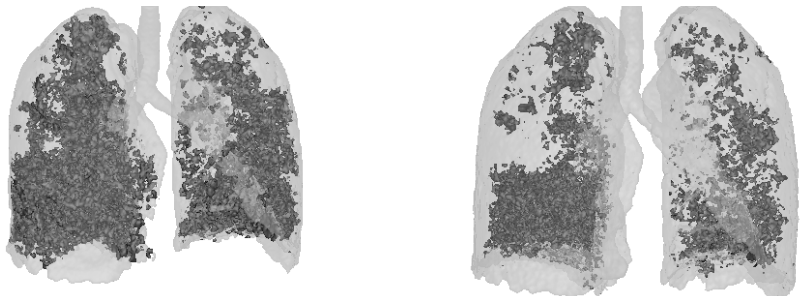


Fig. 3. 3-D volume renderings for patient D show the distribution of alveolar disease pre- (left) and post-lavage (right). Note the regions of increased aeration, particularly in the upper lobes, also reflected by the shift of parenchymal intensities towards −1000 HU, or better ventilated lung.

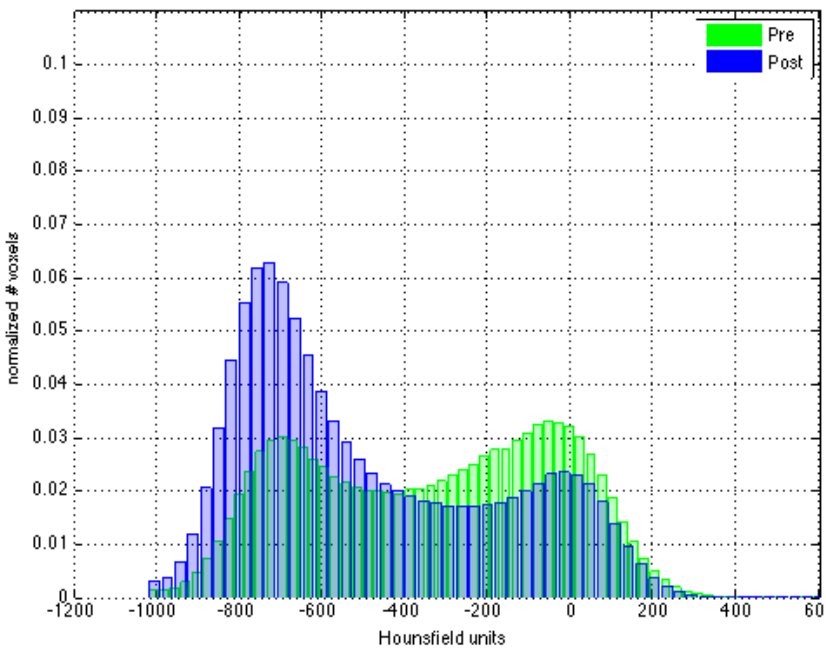


Fig. 4. Normalized histogram analysis of both lungs in patient D pre- and post-WLL demonstrates a leftward shift of parenchymal intensities (toward -1000 HU, the attenuation of air) as alveolar deposits are removed from the lungs by lavage. In addition, patients may be able to sustain deeper inspirations during their CT studies as a result of treatment.

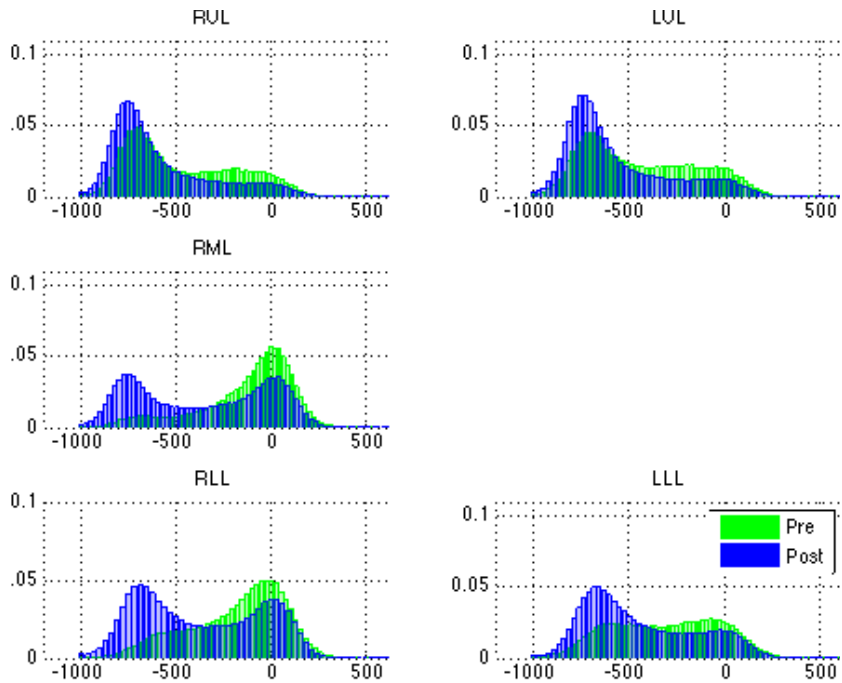


Fig. 5. Normalized lobar histogram analysis in patient D pre- and post-WLL similarly demonstrates a leftward shift of parenchymal intensities as alveolar deposits are washed from the lungs by lavage. Lobar identifications are described in table 1.

characterize the effects of WLL in five patients with PAP. We are able to regionally quantify the percentage of lung volume that is "recovered", i.e. depleted of excess surfactant, after saline lavage. This method has great potential clinical utility. It could assist pulmonologists in determining not only whether lavage has been effective, but also in regionally quantifying disease improvement as well as pinpointing areas of lung that are refractory to treatment and adjusting subsequent treatments accordingly. This methodology can further be applied in the realm of drug therapy for PAP, in order to evaluate the efficacy of pharmacologic agents under development and eventually to monitor patients' progress during treatment.

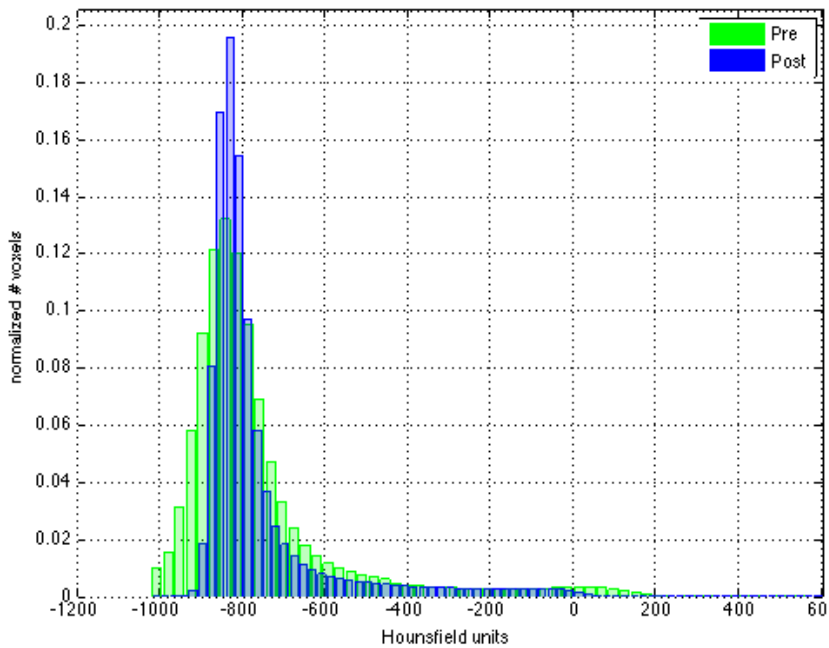


Fig. 6. Normalized histogram analysis of both lungs in patient C pre- and post-WLL does not reveal much response to treatment as compared to patient D in figure 4. Comparison of these normalized histograms can be used to assess treatment response within and across individuals.

However, there are limitations of this work that need to be addressed before this analysis can become an effective complement to patient care. In patients with widespread disease that involved sub-pleural lung (e.g., patient D), the initial automated segmentation failed to detect lung in regions of disease adjacent to the liver and chest wall. An improved initial segmentation that incorporates both intensity- and edge-based criteria would probably improve this step. Further-

more, automated lobar segmentation, instead of our semi-automated approach, would make the analysis timeline more realistically applicable to the clinical setting. However, the same challenge to the whole-lung segmentation also applies to fissure detection, and more work needs to be done to develop methods that can process this type of difficult data.

In preliminary experiments, we anticipated that a single range of intensities could be used to segment diseased lung in all patients with PAP. However, given the vast spectrum of disease and the combination of alveolar and interstitial patterns of disease that can be seen on CT with PAP, it became clear that thresholds needed to be customized to the individual patient. Future work will involve histogram manipulation for disease detection, so that intensity thresholds can be automatically extracted from the data to reflect the inherent severity of disease in the patient. In addition, we intend to validate our disease segmentation methodology using comparisons to manual segmentations of areas of disease performed by expert radiologists.

Lung volumes in the pre- and post-lavage studies are inherently different as patients are imaged at different times and may not breath-hold at the same volumes. In future analyses, we plan to incorporate registration of pre- and post-lavage datasets to eliminate this variable and generate a more accurate analysis of treatment efficacy.

This methodology can be applied to assess disease progression and response to treatment based on imaging findings in many pulmonary pathologies. These techniques would be particularly useful for infiltrative diseases such as idiopathic pulmonary fibrosis or a variety of chronic interstitial pneumonias that manifest with both interstitial and alveolar components. Furthermore, this work is an initial step towards enabling more quantitative reporting of clinical studies in radiology.

References

1. Trapnell, B.C., Whitsett, J.A., Nakata, K.: Pulmonary alveolar proteinosis. NEJM **249**(26) (2003) 2527–2539
2. Ioachimescu, O.C., Kavuru, M.S.: Pulmonary alveolar proteinosis. Chronic Respiratory Disease **3**(3) (2006) 149–159
3. Beccaria, M., Luisetti, M., Rodi, G., Corsico, A., Zoia, M., Colato, S., Pochetti, P., Braschi, A., Pozzi, E., Cerveri, I.: Long-term durable benefit after whole lung lavage in pulmonary alveolar proteinosis. Eur Respir J **23**(4) (2004) 526–531
4. Lee, C.H.: The Crazy-paving Sign. Radiology **243**(3) (2007) 905–906
5. Holbert, J.M., Costello, P., Li, W., Hoffman, R.M., Rogers, R.M.: CT Features of Pulmonary Alveolar Proteinosis. Am. J. Roentgenol. **176**(5) (2001) 1287–1294
6. National Library of Medicine http://www.itk.org: Insight Segmentation and Registration Toolkit. (2003)
7. Hu, S., Hoffman, E., Reinhardt, J.: Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. IEEE TMI **20**(6) (2001) 490–498
8. Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., s. Ho, Gee, J.C., Gerig, G.: User-guided 3d active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. Neuroimage (2006)

9. ANTS: Advanced normalization tools. `http://sourceforge.net/projects/advants`
10. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med Image Anal **12**(1) (Feb 2008) 26–41

Feature Analysis of Hyperpolarized Helium-3 Pulmonary MRI in Asthmatics versus Non-Asthmatics

N. J. Tustison[1], T. A. Altes[2], G. Song[1], J. P. Mugler III[2], E. E. de Lange[2], and J. C. Gee[1]

[1] Penn Image Computing and Science Laboratory, University of Pennsylvania, Philadelphia, PA, USA
[2] Department of Radiology, University of Virginia, Charlottesville, VA, USA

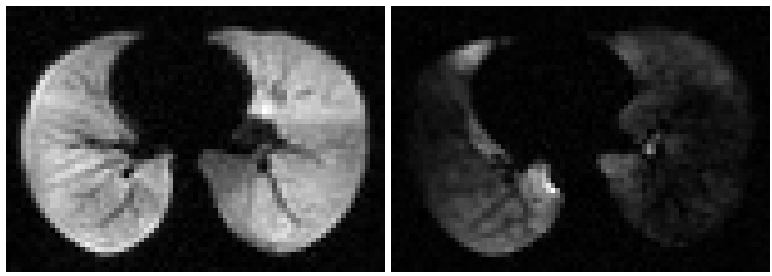
Abstract. Computational techniques for parenchymal characterization using CT have demonstrated significant research potential [1, 2]. We describe similar quantitative analysis of the lung using hyperpolarized helium-3 MR ventilation image data. This study consisted of a total of 55 subjects (47 asthmatics and 8 non-asthmatics). Each subject was imaged both before and after respiratory challenge. Additionally, each subject underwent a complete pulmonary function exam. Following image acquisition, approximately 1600 statistical features were calculated from the segmented lungs (and sub-regions) in each image. Each of these features were ranked along with the 27 pulmonary function test (PFT) values using a mutual information based feature subset selection algorithm. It is shown that several image features perform much better in characterizing clinical diagnosis compared with the current clinical gold-standard PFT values.

1 Introduction

Recent developments in MRI research utilizing noble gases, such as helium-3 and xenon-129, have demonstrated the capability of visualizing alveolar and bronchial air spaces [3]. Currently, hyperpolarized helium-3 MRI is a low-risk investigatory technique which provides high spatial and temporal resolution images of the air spaces of the lungs.

Ventilation or spin density images are acquired by measuring the signal produced by the helium-3 atoms within each voxel. If the signal intensity in the MR images were solely dependent upon the density of helium-3 atoms in each pixel, these images would directly reflect regional ventilation. However, this is not the case since the coil transmit and receive sensitivity, and the regional partial pressure of oxygen within the lung contribute to the measured signal intensity. For this reason, the helium-3 images provide information about the homogeneity of ventilation within the lung but do not provide a quantitative measure of absolute regional ventilation. When a subject inhales the helium-3 gas, areas of the lung that are well ventilated receive a large volume of helium-3 gas which produces

a strong MR signal causing these areas to appear bright on the resulting MR images [3]. In contrast, regions that are poorly ventilated receive little helium-3 and thus produce a weak MR signal causing these regions to appear dark in the resulting images (Fig. 1).



(a) (b)

Fig. 1. Axial hyperpolarized helium-3 MR ventilation images contrasting well-ventilated lungs (left) with poorly ventilated lungs (right).

Previous methodologies for evaluation of helium-3 ventilation images have been limited to radiological assessment of signal heterogeneity as well as various forms of quantifying *ventilation defects*, i.e. regions of poor ventilation, whether it be the number of such defects or the total volume of such defects. Using hyperpolarized helium-3 MRI, it has been shown that asthmatics have an increased volume of lung regions that are poorly ventilated (ventilation defects) than age-matched normal subjects, and that these defects increase in number with increasing asthma severity or with provocation such as exercise or methacholine [4]. Thus, the ventilation defects on hyperpolarized gas MRI appear to be depicting the reversible airway obstruction that was known to occur in asthmatics but was previously difficult to visualize [5].

In this paper, we describe an automated computational framework for generating and analyzing features from helium-3 ventilation images. Such features attempt to characterize lung parenchyma in the varying stages of pathogenesis. Our computational approach permits the analysis of large studies due to complete automation, is extendible to incorporating new features, and facilitates quick processing of data. Furthermore, we demonstrate that our analysis compares favorably with the gold-standard of pulmonary function testing for characterizing clinical diagnosis of asthmatics.

2 Method

Image processing for each helium-3 image volume requires the following steps:

1. bias field correction,

2. segmentation of the lungs,
3. generation of regional masks,
4. generation of feature images,
5. calculation of the image features, and
6. ranking of the image features in conjunction with the PFT values.

Due to the large quantity of current image data used for the study in addition to future stuides, a software pipeline was developed for automated analysis of helium-3 lung MRI. This pipeline was implemented on a computing cluster that allows for distributed parallel processing of the images. In addition, much of the software development constituting our analysis pipeline has been made publicly available through the Insight Toolkit (ITK) of the National Institutes of Health, an open-source repository for popular image analysis algorithms [6].

2.1 Retrospective Bias Field Correction

Significant bias field effects can be seen in typical helium-3 images. This is manifested as a low frequency intensity artifact across the image. The left column of Figure 2 illustrates this bias field artifact in two subjects. Several algorithms exist for correcting the nonparametric nonuniform intensity in magnetic resonance images caused by field inhomogeneities. One popular algorithm is the non-uniform intensity normalization (N3) approach [7, 8]. A particularly advantageous aspect of this algorithm is that it does not require a prior tissue model for its application. This algorithm was used to estimate the bias field and correct the images shown respectively in the middle and right columns of Figure 2.

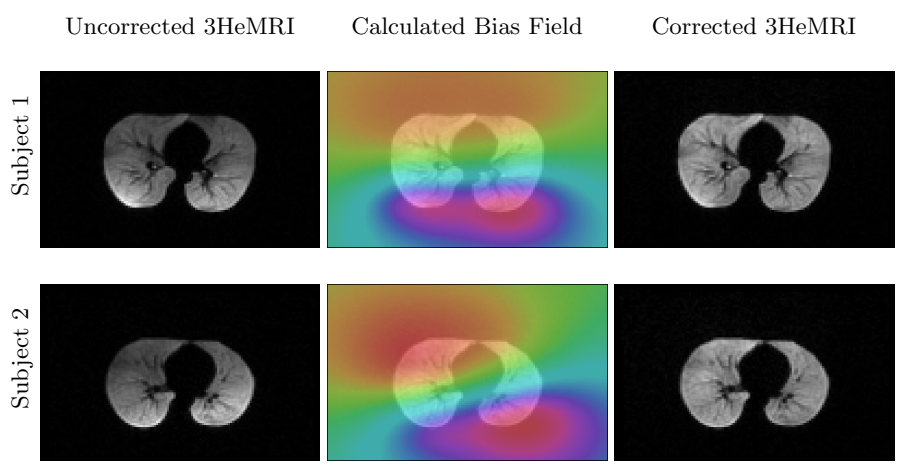
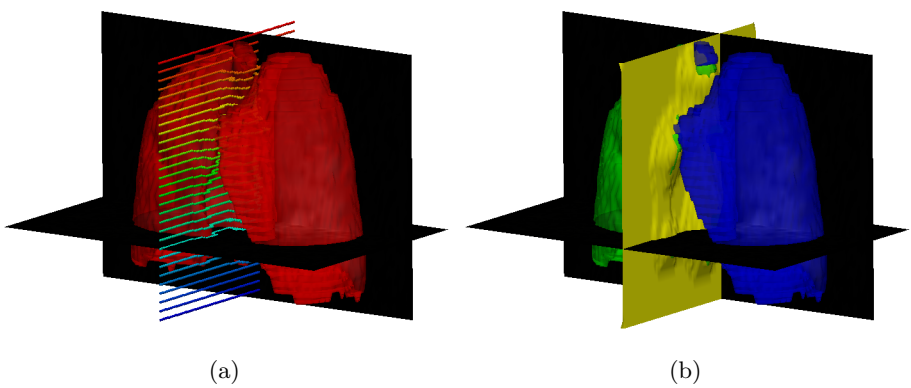


Fig. 2. Left column: Axial 3HeMRI from two subjects evidencing severe bias field artifacts. Middle column: The calculated bias field. Right column: Corrected images.

2.2 Lung Segmentation

An essential precursor to calculation of meaningful image measures is the segmentation of whole lungs from the helium-3 images. Because the quantity of data processed prohibits the use of routine manual segmentation or supervision-intensive semi-automated segmentation methods, we developed an automated segmentation routine specifically tailored for helium-3 lung images.

We first preprocess the images by applying an anisotropic diffusion operation which smoothes both the background noise and the helium-3 signal internal to the lungs while respecting the lung boundaries. Using the statistics surrounding a seed point placed in the background of the image, we grow this region to the rest of the connected background region while iteratively recalculating regional statistics. This separates the whole lung from the background. A 3-D view of a single segmentation is shown in Fig. 3(b).



(a) (b)

Fig. 3. Automatic labeling of left and right lungs. (a) Segmentation of the lungs from the background is illustrated in red. For each axial slice, we generate a minimum-cost path (the minimum path for each axial slice is rendered a different color). (b) Fitting a B-spline surface to these minimum-cost paths labels the left (blue) and right (green) lungs.

Due to the anatomical proximity of the anterior portions of the left and right lungs as well as the presence of helium-3 in the trachea, the whole lung segmentation might produce a single connected component segmented object. To separate the left and right lungs, we iterate through each axial slice and find a minimum path [9] which separates the image slice into left and right halves while respecting the segmented object. A sample set of axial minimum paths is shown in Fig. 3(a). We then fit a smooth surface [10] to the set of axial minimizing paths which separates the whole lung segmented object into left and right components (Fig. 3(b)).

Note that the images acquired from asthmatics used in this study were diagnosed as mild to moderate. As such, the poorly ventilated regions were sufficiently minimal to allow for the described segmentation approach to work for all images in the study. We are currently pursuing a combined template strategy for more severe cases encountered in future studies.

2.3 Generation of Regional Masks

Not only is the calculation of image features over the whole lung essential for adequate lung characterization but feature generation over anatomically-based regional masks also provides important regional information unavailable to pulmonary function testing. Investigation into the distribution of airway disease in CT, such as emphysema, from the inner parenchymal core to the rind has demonstrated that the distribution of disease is generally more extensive in the core than in the rind region of the lung. Thus, corresponding inner core/outer rind regional masks are generated using binary morphological operations. We also further subdivide the left/right lungs into lower and upper portions for additional regional measurements. A total of 13 regional masks were created from the whole lung segmentations described in the previous section. The axial and coronal views of these regional masks are shown in Fig. 4.

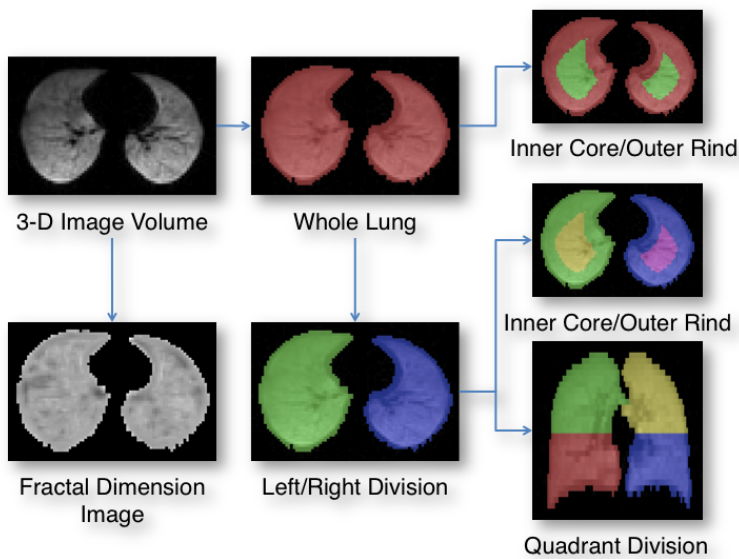


Fig. 4. Flowchart illustrating the generation of both feature images (e.g. stochastic fractal dimension image) and regional masks for localized analysis.

2.4 Image Features

For each of the 13 regional masks, three categories of features were calculated. These feature categories are:

- first order statistics derived from the intensity histogram of the original helium-3 image,
- first order statistics derived from the stochastic fractal dimension image [11], and
- second order statistics, or texture-based measurements, using the cooccurrence [12] and run-length [13] matrices of the original helium-3 image.

Helium-3 Intensity Histogram First Order Statistical Features. Within each of the regional masks, the grey-level intensity histogram is generated. This permits the calculation of the standard first order statistics, i.e. mean, variance, skewness, kurtosis, entropy, $5^{th}\%$ intensity value, mean of $5^{th}\%$, and $95^{th}\%$ intensity value, mean of $95^{th}\%$. From the histogram and image spacing we also calculate the volume of each lung.

Stochastic Fractal Dimension Image Features. Viewed as an intensity surface, Mandelbrot's fractal theory provides an informative framework for characterizing images [11, 14]. It has been shown that first order statistics derived from the stochastic fractal dimension (SFD) image facilitate discriminative feature analysis in CT images [1]. A SFD image is produced by iterating through the original image where, at each voxel, the corresponding fractal value requires inspection of each pair of voxels in the surrounding neighborhood. The average absolute intensity value difference is plotted against the voxel pair wise distances on a log-log scale. The voxel value is equal to 3 minus the slope of the line calculated using linear regression. First-order statistics from the SFD image in the various regions are calculated and used as image features.

Cooccurrence and Run-Length Matrix Features. Second order statistics have demonstrated utility in image texture classification. These include measurements derived from the grey-level cooccurrence matrix (CM) [12] and measurements derived from the run length matrix (RLM) [13]. As mentioned in the introduction, previous methodologies for evaluation of helium-3 MR images have included radiological assessment of "heterogeneity" for which these texture features are a surrogate. The specific set of CM features are energy, entropy, correlation, inverse difference moment, inertia, cluster shade, cluster prominence, and Haralick's correlation. The specific set of RLM features are short run emphasis, long run emphasis, grey level non-uniformity, run-length non-uniformity, run percentage, low grey level run emphasis, high grey level run emphasis, short run low grey level emphasis, short run high grey level emphasis, long run low grey level run emphasis, and long run high grey level run emphasis. This ordering of the CM and RLM features are used as subscripts in subsequent sections.

Otsu Thresholding of Parenchyma. Also utilized in previous research of hyperpolarized helium-3 MRI was radiological assessment of poor ventilation regional quantitation. Since the ventilation images do not reflect absolute regional ventilation we attempted to mimic this measurement in a simple fashion by applying an optimal thresholding to the grey-level histogram of the original image [15] to quantify regions of poor ventilation. We then divide the total volume of the ventilation defect regions by the total volume of the lungs to calculate a ventilation ratio.

2.5 Feature Characterization using Mutual Information

A straightforward approach to finding a subset of features most characterizing of the observed data is finding those features which correlate strongest with the target classification, i.e. the goal of "maximal relevance." The mutual information between two random variables x and y, defined as

$$I(x;y) = \int\int p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dxdy, \tag{1}$$

is also used to define the dependency of variables. Given a set of N samples each characterized by M features $X = \{x_i, i = 1, \ldots, M\}$ where each sample is assumed to correspond to one of a finite number of classes described by the classification vector c, the maximally relevant feature, $x_i \in S \subset X$, maximizes the quantity $D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c)$. In considering multiple features, it is desirable that the selected features be minimally redundant where redundancy is defined as $R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i; x_j)$. The minimal-redundancy-maximal-relevance (mRMR) framework combines these two desiderata for feature subset selection by incrementally maximizing $\Phi = D - R$ [16].

3 Results

Hyperpolarized helium-3 images were acquired from 55 subjects (47 asthmatic and 8 healthy) before and after respiratory provocation (exercise or methacholine-induced). For each of the 110 images, a total of 533 features were calculated over all the anatomic regions. Each subject also underwent pulmonary function testing which generated an additional 27 clinical features, e.g. forced vital capacity (FVC), forced expiratory volume in 1 second (FEV_1), and peak expiratory flow (PEF). For each subject we combined the pre and post respiratory challenge images in addition to their difference values for a total of 1599 features. This was combined with the 27 PFT values for a total of 1626 features per subject.

Two rankings were generated from the mRMR algorithm using mutual information difference. For each entry, we show the rank, the calculated feature, whether the acquisition was pre or post respiratory provocation (or the difference of the two), the region, and the score. Table 1 ranks the relevancy of each feature considered individually whereas the ranking in Table 2 accounts for mutual

redundancy between features. The top 10 individually considered features are given in Table 1. Since these features are strictly image-based, we also situate the top 3 PFT values within the rankings. We do a similar placement for Table 2.

Rank	Feature	Pre/Post/Diff	Region	Score
1	ventilation ratio	pre respiratory challenge	left lung	0.357
2	ventilation ratio	post respiratory challenge	left lung	0.329
3	RLM_5	post respiratory challenge	lower left lung	0.329
4	RLM_{10}	difference	whole lung	0.267
5	RLM_{10}	difference	outer rind, right lung	0.259
6	CM_2	post respiratory challenge	outer rind, left lung	0.198
7	CM_2	post respiratory challenge	lower left lung	0.193
8	RLM_5	post respiratory challenge	left lung	0.182
9	CM_6	post respiratory challenge	left lung	0.180
10	CM_4	post respiratory challenge	left lung	0.169
⋮	⋮	⋮	⋮	⋮
169†	% Predicted FEV_1	—	—	0.082
170†	FEV_3 Final	—	—	0.082
⋮	⋮	⋮	⋮	⋮
182†	% FEV_1	—	—	0.080

Table 1. Maximal relevance ranking (each feature is considered individually) of PFT and imaging features using the mRMR feature classification algorithm. Those rankings marked by a '†' denote a PFT value. The CM and RLM subscripts refer to the respective ordered measurement given in the text. Note the relative performance of the top image-based features compared with the top 3 PFT performers (rank 169, 170, and 182).

4 Discussion

There are several interesting observations to be made from the results. The first is rather obvious in that, considered as a global assessment, individual image features perform much better in characterizing clinical diagnosis compared with the PFT values. Another interesting observation is that despite the relatively poor individual classification performance of the PFT values, the better performers of this group have a minimal redundancy with respect to the image features. This is supported by the correlation values between the PFT results and each of the 1599 image features. For example the measurement % Predicted FEV_1 (rank 7 in Table 2) has correlation values with the image features in the range $[-0.4012, 0.4328]$ demonstrating practically no correlation with any of the image features. This demonstrates the orthogonal nature of the image-based information relative to the information provided by the PFT values.

Rank	Feature	Pre/Post/Diff	Region
1	ventilation ratio	pre respiratory challenge	left lung
2	RLM_5	post respiratory challenge	outer rind, left lung
3	ventilation ratio	post respiratory challenge	left lung
4	RLM_{10}	difference	right lung
5	CM_2	post respiratory challenge	lower left lung
6	RLM_8	post respiratory challenge	lower left lung
7†	% Predicted FEV_1	—	—
8	RLM_5	post respiratory challenge	lower left lung
9	CM_6	post respiratory challenge	outer rind, left lung
10	CM_2	post respiratory challenge	outer rind, whole lung
⋮	⋮	⋮	⋮
34†	FEV_1	—	—

Table 2. Maximal relevance, minimal redundancy ranking of PFT and imaging features using the mRMR feature classification algorithm. Those rankings marked by a '†' denote a PFT value. The CM and RLM subscripts refer to the respective ordered measurement given in the text.

It is also demonstrated that the ventilation ratio was the top classification feature, boding well for previous radiological assessments which attempted a similar calculation in quantifying regions of ventilation defects [4]. Also consistent with clinical understanding is the prominence of "post respiratory challenge" features amongst the top of the rankings since such provocation tends to exacerbate the asthmatic condition which is presumably reflected in the post images.

References

1. Xu, Y., van Beek, E.J.R., Hwanjo, Y., Guo, J., McLennan, G., Hoffman, E.A.: Computer-aided classification of interstitial lung diseases via MDCT: 3D adaptive multiple feature method (3d amfm). Acad Radiol **13**(8) (Aug 2006) 969–978
2. Sorenson, L., Shaker, S.B., de Bruijne, M.: Texture based emphysema quantification in lung ct. In: First International Workshop on Pulmonary Image Processing. (2008) 5–14
3. de Lange, E.E., Mugler, J.P., Brookeman, J.R., Knight-Scott, J., Truwit, J.D., Teates, C.D., Daniel, T.M., Bogorad, P.L., Cates, G.D.: Lung air spaces: Mr imaging evaluation with hyperpolarized 3he gas. Radiology **210**(3) (Mar 1999) 851–857
4. de Lange, E.E., Altes, T.A., Patrie, J.T., Parmar, J., Brookeman, J.R., Mugler, J.P., Platts-Mills, T.A.E.: The variability of regional airflow obstruction within the lungs of patients with asthma: assessment with hyperpolarized helium-3 magnetic resonance imaging. J Allergy Clin Immunol **119**(5) (May 2007) 1072–1078
5. Altes, T.A., Powers, P.L., Knight-Scott, J., Rakes, G., Platts-Mills, T.A., de Lange, E.E., Alford, B.A., Mugler, J.P., Brookeman, J.R.: Hyperpolarized 3he mr lung ventilation imaging in asthmatics: preliminary findings. J Magn Reson Imaging **13**(3) (Mar 2001) 378–384

6. Ibanez, L., Schroeder, W., Ng, L., Cates, J.: The ITK Software Guide. 2 edn. Kitware, Inc., Albany, NY (November 2005)
7. Sled, J.G., Zijdenbos, A.P., Evans, A.C.: A nonparametric method for automatic correction of intensity nonuniformity in mri data. IEEE Trans Med Imaging **17**(1) (Feb 1998) 87–97
8. Tustison, N.J., Gee, J.C.: N4ITK: Nick's N3 ITK implementation for MRI bias field correction. Insight Journal (2009)
9. Barrett, W.A., Mortensen, E.N.: Interactive live-wire boundary extraction. Medical Image Analysis **1**(4) (September 1997) 331–341
10. Tustison, N.J., Gee, J.C.: Generalized n-d C^k B-spline scattered data approximation with confidence values. In: Proc. Third International Workshop Medical Imaging and Augmented Reality. (2006) 76–83
11. Chen, C.C., DaPonte, J., Fox, M.: Fractal feature analysis and classification in medical imaging. IEEE Transactions on Medical Imaging **8**(2) (1989) 133–142
12. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics **3**(6) (1973) 610–621
13. Dasarathy, B.R., Holder, E.B.: Image characterizations based on joint gray-level run-length distributions. Pattern Recognition Letters **12** (1991) 497–502
14. Tustison, N.J., Gee, J.C.: Stochastic fractal dimension image. Insight Journal (2009) published online
15. Otsu, N.: A thresholding selection method from gray-scale histogram. IEEE Transactions on System, Man, and Cybernetics **9** (1979) 62–66
16. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. Pattern Anal. Mach. Intell. **27**(8) (2005) 1226–1238

Shielding Active Shape Models against Weak Lung Field Boundaries for Segmentation of Chest Radiographs

Michalis A. Savelonas, Dimitris K. Iakovidis

Dept. of Informatics and Comp. Technology, Technol. Educ. Institute of Lamia, Greece

Abstract. A novel active shape model (ASM) segmentation scheme is proposed, for the detection of the lung field boundaries in chest radiographs. The proposed scheme is robust in the presence of weak lung field boundaries, which are recognized as a common cause of missegmentation. This situation is prevalent in chest radiographs obtained from patients with abnormalities, such as lung consolidations, or even in normal chest radiographs of low quality. In order to shield the ASM against the presence of weak boundaries, the distance used in searching for the optimal displacements of the ASM landmarks is modified so that it incorporates both shape prior and patient-specific anatomical information. The segmentation performance of the proposed scheme is experimentally validated on a challenging set of chest radiographs obtained from patients with bacterial pulmonary infections. The results obtained indicate its advantage over the original ASM.

1 Introduction

A significant research effort has been invested in the segmentation of chest radiographs [1], including rule-based methodologies [2], neural network, Markov random fields [3] and other pattern classification approaches [4], active contour models [5], active shape models [6],[7] and graph cuts [8]. However, accurate segmentation of lung field boundaries defined by weak edges remains an open issue. Such weak edges are associated with the presence of abnormalities or with poor imaging quality and can be hardly distinguished even by experienced physicians [9].

Recently, a rule-based selective thresholding approach coping with the issue of weak edges has been proposed in [10]. This approach is unsupervised and defines a set of salient control points around the lung fields. The salient point detection process is supported by a selective thresholding algorithm that cuts off image intensities based on local intensity histograms sampled from the spinal cord. The detected points are then intuitively interpolated by Bézier curves [11]. The preliminary results presented indicate its robustness against the presence of weak edges, as well as its advantage against the graph cuts approach [8]. A drawback of this methodology is that it may produce implausible shapes, as a result of the shape variability characterizing lung field boundaries in chest radiographs.

Among the various lung field segmentation methodologies, the shape-constrained deformability of ASMs makes them a competent candidate for coping with the shape variability of the lung fields. According to the original ASM approach [12] the shape

of an image region is represented by the principal components of landmark point vectors, whereas the grey-level appearance of that region is limited to its border. The latter consists of the normalized first derivative of the grey-level intensity profiles centered at each landmark that run perpendicular to the region's contour. This formulation makes it evident that ASMs are quite sensitive to weak lung field boundaries, which affect the derivatives of the intensity profiles.

In this paper, a novel ASM-based segmentation scheme is proposed for the detection of lung field boundaries in chest radiographs, which copes with the presence of weak edges. The proposed scheme introduces a modified distance measure guiding the evolution of the ASM, which utilizes patient-specific anatomical information derived from the rule-based selective thresholding approach [10], so as to provide robustness against the presence of weak lung field boundaries. The ASM formulation of the proposed scheme accounts for the shape variability of the lung fields. The segmentation performance of the proposed scheme is experimentally validated on a challenging set of chest radiographs obtained from patients with bacterial pulmonary infections.

The rest of this paper consists of four sections. In section 2 a brief description of the utilized rule-based selective thresholding approach is provided. The modified ASM approach is presented in Section 3. In Section 4 the results of its experimental evaluation are presented, whereas in the last section the conclusions of this study are summarized.

2 Rule-Based Selective Thresholding

Let I be a new chest radiograph of size $N \times M$ pixels, which is uniformly sampled from top to bottom with s_h non-overlapping rectangular windows of $h \times M$, pixels, where $h<N$. For each sample an average horizontal profile of grey-level intensities is estimated. For each profile a local maximum is selected as a spinal cord point according to [10]. The selective thresholding algorithm proceeds as follows:

Step 1. Sample the radiograph across spinal cord:
- For each of the s_h points detected on the spinal cord:
  - Acquire a square sample of x^2 pixels.
  - For each sample i=1,…, s_h:
    - Calculate its intensity histogram h_i,
    - Select a set of histogram components ph_i centered at its highest peak.

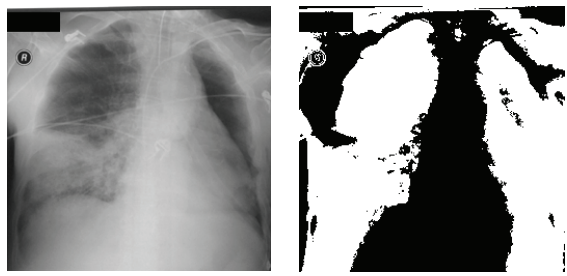
Step 2. Accumulate the values ph_i from all samples into a single histogram H.

Step 3. Find the last non-zero component m of H.

Step 4. Generate an output image $T(I)$ from I as follows:
- Set the intensities of I that correspond to the non-zero components of H, to zero.
- Set the intensities of I that are larger than m, to zero.
- Set all non-zero intensities to one to obtain a binary image.

The intensities that are larger than m are subtracted from I because they are unlikely

to belong to the lung fields since the spinal cord is generally brighter than both normal lung parenchyma and consolidations [13]. Such regions may include dense objects used for patient's monitoring or support. Figure 1 illustrates an example of a chest radiograph as well as the resulting thresholded image.



(a) (b)

Fig. 1. Example image. (a) Radiograph *I*, (b) Thresholded image *T*(*I*).

3 ASM Evolution guided by a Modified Distance Measure

The ASM approach [12] captures shape variability of the lung fields by applying principal component analysis (PCA) to build a shape model from a set of training shape vectors, which encode ground truth lung field boundaries. Shape vectors are defined as $\mathbf{x}=(x_1,y_1,\ldots,x_n,y_n)^{\mathrm{T}}$, where $(x_1,y_1),\ldots,(x_n,y_n)$ are n landmark points. The shape model is represented by the covariance matrix of the shape vectors $\mathbf{C}_{\mathbf{x}}$ from which t eigenvectors ϕ_i, $i=1,\ldots,t$ corresponding to the largest eigenvalues λ_i, are estimated.

The value of t is determined as the smallest t for which the following inequality holds [14]:

$$\sum_{i=1}^{t} \lambda_i \geq f_v \sum_{i=1}^{2n} \lambda_i \qquad (1)$$

where f_v is the desired percentage of shape variation to be represented by the t principal eigenvectors. The eigenvectors form the matrix $\Phi = (\phi_1 | \phi_2 | \ldots | \phi_t)$ which is used for the approximation of a shape $\mathbf{x}$ in

$$\mathbf{x} = \bar{\mathbf{x}} + \Phi \cdot \mathbf{b}_{\mathbf{x}} \qquad (2)$$

where $\bar{\mathbf{x}}$ denotes the mean of N training shape vectors, and $\mathbf{b}_{\mathbf{x}}$ holds the shape parameters, estimated by $\mathbf{b}_{\mathbf{x}} = \Phi^{\mathrm{T}}(\mathbf{x} - \bar{\mathbf{x}})$.

In addition to the shape, the ASM models the local appearance of the image region around each landmark. The local appearance model is represented by the mean $\bar{\mathbf{d}}_i$ and the covariance matrix $\mathbf{C}_{\mathbf{d}i}$ of the normalized first derivatives $\mathbf{d}_i = (d_{i1}, d_{i2}, \ldots, d_{i2k})^{\mathrm{T}}$, $k>0$, of image intensity profiles $\mathbf{g}_i = (g_{i1}, g_{i2}, \ldots, g_{i2k+1})^{\mathrm{T}}$. The

latter are sampled around each landmark i=1,…,n over the N training shapes. The trained ASM is then applied on a new chest radiograph and evolves to detect the lung field boundaries.

The ASM evolves according to a simple iterative scheme [11]. Each landmark can move along a direction perpendicular to the contour by s positions on either side of the contour, evaluating 2s+1 positions, in total. In the original ASM formulation, the new position (x'_i, y'_i) of each landmark i=1,…,n, on the search direction, is determined as the one minimizing the Mahalanobis distance $\mu(\mathbf{d}'_{ij})$ between the vector of the normalized first derivatives estimated from the 2s+1 pixel intensity profile centered at each of the possible new landmark positions (x'_{ij}, y'_{ij}), i=1,…,n, j=1,…,2s+1, and $\bar{\mathbf{d}}_i$:

$$\mu(\mathbf{d}'_{ij}) = (\mathbf{d}'_{ij} - \bar{\mathbf{d}}_i)^{\mathrm{T}} \mathbf{C}_{\mathbf{d}i}^{-1} (\mathbf{d}'_{ij} - \bar{\mathbf{d}}_i) \quad (3)$$

After all the landmarks are updated, the shape model is fitted to the new landmarks. It is evident that the Mahalanobis distance is sensitive to weak lung field boundaries, since the derivatives of the intensity profiles are affected.

The proposed ASM scheme utilizes a modified distance measure μ_a, which incorporates intensity gradient prior information quantified by Mahalanobis distance μ, as well as patient-specific anatomical information derived from the thresholded image $T(I)$ generated by the rule-based selective thresholding approach described in Section 2. The modified distance μ_a is defined as follows:

$$\mu_a = \mu + \lambda d \quad (4)$$

where λ is a weight parameter and d is the Euclidean distance between each of the possible new landmark positions (x'_{ij}, y'_{ij}) and the transition point (x_i^{tr}, y_i^{tr}), i=1,…,n, j=1,…,2s+1. The latter is defined for each landmark i, as the closest point to i along the direction perpendicular to the contour, which is associated with an intensity transition on the thresholded image $T(I)$.

Intuitively, Eq. (4) is based on the consideration that in cases of weak lung field boundaries μ is expected to be less discriminative due to its dependence on the normalized first derivatives. However, among the new candidate landmark positions, the one associated with the smaller Euclidean distance d, is more likely to correspond to the actual lung field boundaries. Thus, the patient-specific anatomical information provided by $T(I)$, complements the intensity gradient prior information quantified by μ and allows μ_a to provide a more complete measure of the likelihood of a candidate landmark position to correspond to the actual lung field boundaries.

The incorporation of a multiresolution local appearance model helps the algorithm to avoid trapping into irrelevant regions, whereas the intensity gradient prior information included in the ASM contributes to the plausibility of the output lung field shapes.

The algorithm of the proposed ASM scheme proceeds as follows:

Step 1. Update landmark positions:

- For each possible new landmark position (x'_{ij}, y'_{ij}), i=1,…,n, j=1,…,2s+1:
  - Calculate a profile of normalized first derivatives

$\mathbf{d}'_{ij} = (d_{ij1}, d_{ij2}, ..., d_{ij2k})^{\mathrm{T}}$ on I, centered at this landmark position perpendicular to the contour.

- Calculate the modified distance $\mu_a(\mathrm{d}'_{ij})$.
- Find $\mu_{\min}(\mathrm{d}'_i) = \min_{j=1,...,2s+1}\left(\mu_a(\mathrm{d}'_{ij})\right)$.
- Set $(x'_i, y'_i) = (x'_{ij}, y'_{ij})$ as the new position of landmark i.

Step 2. Fit the current shape model to the updated landmarks:

- Set $\mathbf{x}' = (x'_1, y'_1, ..., x'_n, y'_n)^{\mathrm{T}}$.
- Calculate the new model parameters for $\mathbf{x}'$ by solving Eq. (2) for $\mathbf{b}_{\mathbf{x}'}$.
- Constrain each component $|b_l| < c\sqrt{\lambda_l}$ of $\mathbf{b}_{\mathbf{x}'}$, c>0, l=1,…t, to ensure plausible shapes.

Step 3. Repeat steps 1 to 3 until a proportion p_{close} of points ends up within s/2 of its previous position, or $M_{\max}$ iterations are reached.

Step 4. Move to the next resolution level.

Step 5. Repeat steps 1 to 4 until the finest resolution level $R_{\max}$ is reached.

4 Results

The experimental evaluation of the proposed approach was performed on a set of 96 anonymous chest radiographs obtained with a portable x-ray device from 24 critically ill patients with pulmonary bacterial infections manifested as consolidations. For each patient, a sequence of four chest radiographs was obtained, depicting the progress of the infection at different stages.

Each sequence used in the experiments was obtained after the results of microbiological culture had been issued and the appropriate treatment had been prescribed. As a result of the treatment, in all 24 used sequences the infection diminished with time, the consolidations gradually disappeared and the edge strength associated with lung field boundaries increased. Figure 2 illustrates an example of such a sequence.

All radiographs used in the experiments where digitized at 8 bits and have been downscaled to fit a 256×256-pixel bounding box. Initial pilot experiments which were performed on the complete dataset of 96 images revealed that the image downsampling does not significantly affect the obtained segmentation performance, whereas it reduces the execution time of the algorithm. In addition, these pilot experiments allowed the selection of a fixed shape model parameter setting that yielded good performance. A shape model explaining 98% of the variance (f_v=0.98) was constructed based on the application of PCA on the set of 96 training shape vectors associated with ground truth lung field boundaries, as described in Section 3.

Other settings include three levels of resolution (R_{max}=3), thirty iterations per resolution level (M_{max}=30) and search regions associated with profiles of length five (s=2). When fitting the shape model to the displaced landmarks, each mode was constrained within two times (c=2) the standard deviation. The weight parameter was set equal to one (λ=1), considering the results derived for λ=0.5, 1, 1.5, which indicate

that the influence of the two terms in (4) is approximately equal. Additional values of parameters associated with the sizes used for the salient point detection and selective thresholding include h=9 and x=32.

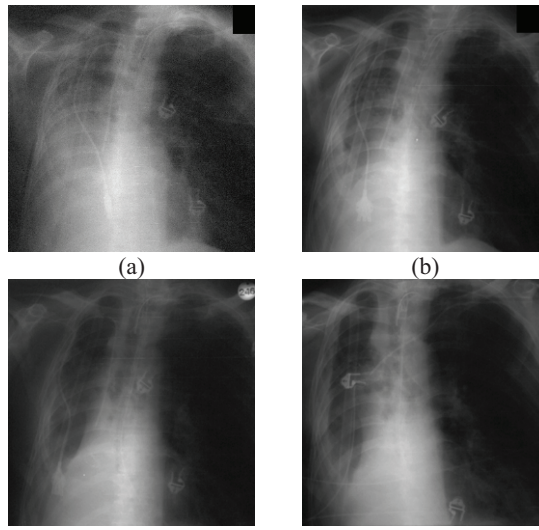


Fig. 2. Example sequence of 4 chest radiographs of the same patient, as the infection of the left lung is diminishing with time.

The performance of the proposed methodology was assessed in terms of Ω, a widely accepted measure of accuracy for binary segmentation tasks:

$$\Omega = \frac{TP}{TP + FP + FN} \quad (5)$$

where *TP* stands for true positive (the area classified as lung by both the proposed methodology and the expert), *FP* for false positive (area classified as background by the expert and as lung by the proposed methodology) and *FN* for false negative (area classified as lung by the expert and as background by the proposed methodology). This measure is a more suitable measure of segmentation performance than the accuracy used in [10], since the latter counts *TN* pixels as correctly detected, providing deceptively high results in cases of relatively small target objects.

A cross-validation scheme involving a total of 24 experiments was employed for the evaluation of the proposed scheme and the original ASM. In each experiment, 92 images comprising of 23 sequences of 4 images each, were used for training, whereas the remaining 4 images of the dataset, which did not belong to the training set and were associated with the same sequence, were used for testing.

Figure 3 presents the progress in time of the mean Ω, obtained as the infection was diminishing, by using the proposed scheme and the original ASM. The value i of the horizontal axis corresponds to the i-th images of all 24 time-sorted sequences, i=1,…,4. Considering that the infections were diminishing with time for all sequences

used, as i increases the mean Ω is calculated over images of increasing density of consolidations. The results indicate that: a) the proposed scheme achieves higher Ω than the original ASM, b) mean Ω obtained by the proposed scheme exceeds 0.85 for all levels of consolidation densities, c) the difference of mean Ω obtained by the proposed scheme and the original ASM is higher in the first images of the sequences, were the density of consolidations is higher.

Figure 4 illustrates the segmentation results obtained by the application of the proposed scheme and the original ASM on the sequence of Fig. 2. It can be observed that unlike the original ASM, the proposed segmentation scheme avoids contour leaking, even in cases of lung field boundaries characterized by weak edges (Fig, 4a1).

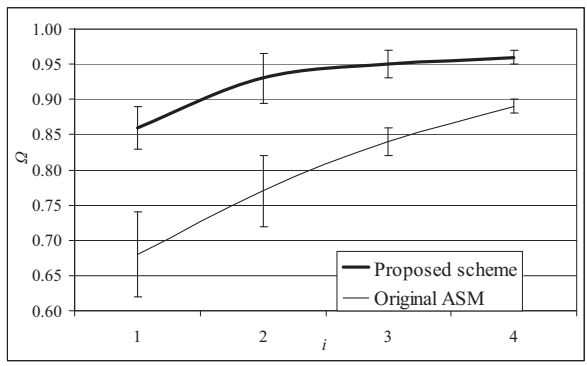


Fig. 3. Progress in time of the mean Ω, obtained as the infection was diminishing.

5 Conclusion

We presented a novel ASM-based segmentation scheme for robust detection of the lung field boundaries in chest radiographs. The proposed scheme utilizes a modified distance measure that incorporates both intensity gradient prior and patient-specific anatomical information to obtain insensitivity to the presence of weak boundaries. The results of its experimental evaluation lead to the following conclusions:

- It achieves more accurate detection of the lung field boundaries than the original ASM.
- It provides robust segmentation of the lung fields in the presence of weak boundaries.
- Its advantageous segmentation performance becomes more evident as the lung field boundaries become weaker.

The proposed scheme introduces a framework, which could potentially embed alternative feature vectors, such as the feature vector calculated by the application of the modified scale invariant feature transform (SIFT). SIFT is supported by recent studies [15] as more distinctive than the general intensity and gradient features. Future work also includes a thorough investigation of the impact of various parameter values on the results and the integration of the proposed lung field detection scheme into a multimodal data mining system for adverse events detection, which will be

capable of co-evaluating radiographic findings of patients with bacterial infections [16].

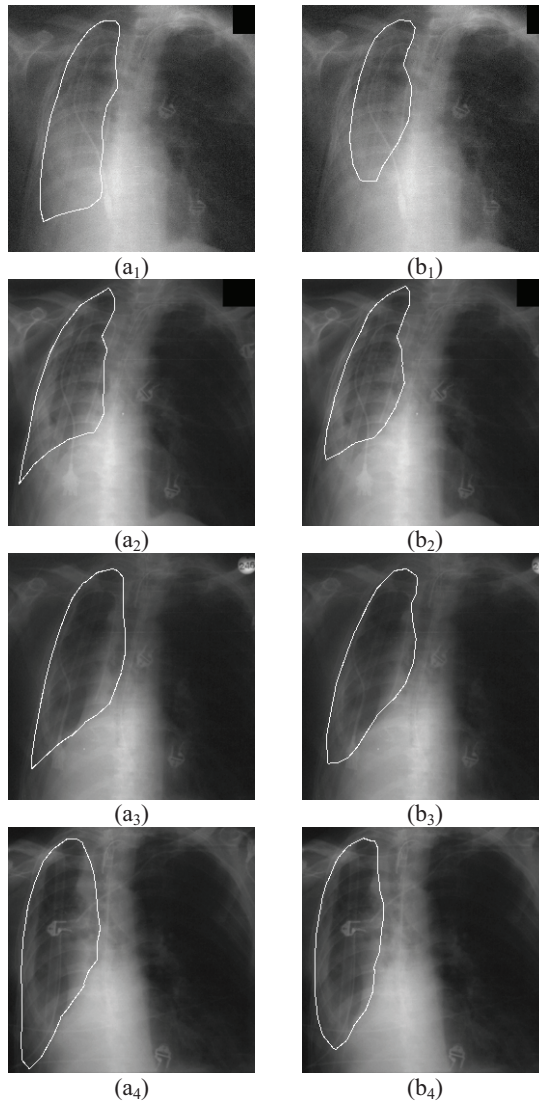


Fig. 4. Segmentation results obtained by the application of the proposed scheme (a_1-a_4) and the original ASM (b_1-b_4) on the sequence of Fig. 2.

Acknowledgement

Great thanks to Dr. G. Papamichalis, M.D. who generously offered his help and advice on the medical aspects of this study. The radiographs have been provided in

anonymized form by the General Hospital of Athens “Sotiria” for the purposes of the project DEBUGIT. This work was supported in part by the European Commission’s Seventh Framework Information Society Technologies (IST) Programme, Unit ICT for Health, project DEBUGIT (no. 217139).

References

1. Ginneken B.V., Romeny B.T.H., Viergever M.A.: Computer-Aided Diagnosis in Chest Radiography: A Survey. IEEE Trans. Medical Imaging, Vol. 20, No. 12 (2001) 1228-1241
2. Ginneken B.V., Romeny B.T.H.: Automatic Segmentation of Lung Fields in Chest Radiographs. Medical Physics, Vol. 27, No. 10 (2000) 2445-2455
3. McNitt-Gray M.F., Huang H.J., Sayre J.W.: Feature Selection in the Pattern Classification Problem of Digital Chest Radiograph Segmentation. IEEE Trans. Medical Imaging, Vol. 14, No. 3 (1995) 537–547
4. Vittitoe N., Vargas-Voracek R., Floyd C. Jr: Markov Random Field Modeling in Posteroanterior Chest Radiograph Segmentation. Medical Physics, Vol. 26, No. 8 (1999) 1670-1677
5. Coppini G., Diciotti S., Falchini M., Villari N., Valli G.: Neural Networks for Computer-Aided Diagnosis: Detection of Lung Nodules in Chest Radiograms. IEEE Trans. Information Technology in Biomedicine, Vol. 7, No. 4 (2003) 344-357
6. Ginneken B.V., Stegmann M.B., Loog M.: Segmentation of Anatomical Structures in Chest Radiographs using Supervised Methods: A Comparative Study on a Public Database. Medical Image Analysis, Vol. 10 (2006) 19-40
7. Ginneken B.V., Frangi A.F., Staal J.J., Romeny B.T.H., Viergever M.A.: Active Shape Model Segmentation with Optimal Features. IEEE Trans. Medical Imaging, Vol. 21 (2002) 924-933
8. Chen S., Cao L., Liu J., Tang X.: Automatic Segmentation of Lung Fields from Radiographic Images of SARS Patients Using a New Graph Cuts Algorithm. Proc. Int. Conf. Pattern Recognition (ICPR), Vol. 1 (2006) 271-274
9. Müller N.L., Franquet T., Lee K.S., Isabela C.: Silva S.: Imaging of Pulmonary Infections, Lip. Williams & Wilkins (2006)
10. Iakovidis D.K., Papamichalis G.: Automatic Segmentation of the Lung Fields in Portable Chest Radiographs Based on Bézier Interpolation of Salient Control Points. Proc. IEEE Int. Conf. Imag. Sys. Tech., Greece (2008) 82-87
11. Bartels R.H., Beatty J.C., Barsky B.A.: An Introduction to Splines for Use in Computer Graphics and Geometric Modelling. San Francisco, CA: Morgan Kaufmann (1998) 211-245
12. Cootes T., Taylor C., Cooper D., Graham J.: Active Shape Models - Their Training and Application. Computer Vision and Image Understanding, Vol. 61, No. 1 (1995) 38-59
13. Novelline R.A.: Squires’s Fundamentals of Radiology. Cambridge: Harvard University Press (1997)
14. Ginneken B.V., Katsuragawa S., Romeny B.T.H., Doi K., Viergever M.A.: Automatic Detection of Abnormalities in Chest Radiographs Using Local Texture Analysis. IEEE Trans. Medical Imaging, Vol. 21, No. 2 (2002) 139-149
15. Shi Y., Qi F., Xue Z., Chen L., Ito K., Matsuo H., Shen D.: Segmenting Lung Fields in Serial Chest Radiographs Using Both Population-Based and Patient-Specific Shape Statistics. IEEE Trans. Medical Imaging, Vol. 27, No. 4 (2008) 481-494
16. Lovis C., Colaert D., Stroetmann V.N.: DebugIT for Patient Safety-Improving the Treatment with Antibiotics through Multimedia Data Mining of Heterogeneous Clinical Data, Stud. Health Technol. Inform., Vol. 136 (2008) 641-646

A Comparative Study of HRCT Image Metrics and PFT Values for Characterization of ILD and COPD

Gang Song[1], Nicholas Tustison[1], Eduardo Barbosa Jr[2], James C. Gee[1], Warren Gefter[2], Kreider Maryl[3], Drew Torigian[2]

[1] Penn Image Computing and Science Laboratory (PICSL), Dept. of of Radiology, University of Pennsylvania, Philadelphia, PA, USA,
[2] Dept. of Radiology, University of Pennsylvania, Philadelphia, PA, USA
[3] Dept. of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Abstract. Several image metrics have been proposed for pulmonary assessment via thoracic high-resolution computed tomography (HRCT) for various pathologies. This paper describes a systematic analysis of the utility of such metrics for characterizing interstitial lung disease (ILD) and chronic obstructive pulmonary disease (COPD), in comparison to data from pulmonary function testing (PFT). HRCT inspiratory and expiratory images for 14 patients with ILD and 11 patients with COPD were acquired retrospectively. PFT values were also acquired retrospectively for each patient. Using a statistical feature selection scheme, our study demonstrates that the quantitative image features perform quite well in comparison with the clinically-used PFT values. In the first 25 selected features out of the total 114 mixed image metrics and PFT values, 21 are from the image metrics. The classification using mixed selected image features and PFT values outperforms using PFT values alone. Our study also shows that these image metrics are not redundant with respect to the PFT values for characterization of ILD and COPD.

1 Introduction

Chronic lung disease constitutes a major worldwide public health care problem and is the fourth leading cause of morbidity and mortality in the United States. Based on clinical, imaging and pathological characteristics, most can be grouped within two basic categories: interstitial lung disease (ILD) and chronic obstructive pulmonary disease (COPD). ILD is a heterogeneous group of diseases in which the hallmark is chronic, progressive, predominantly interstitial inflammation with varying degrees of fibrosis of the lung parenchyma, eventually leading to reduced lung volume, decreased lung compliance and restrictive physiology. COPD is characterized by chronic airflow limitation due to small airway disease and parenchymal destruction which is not fully reversible and is usually progressive.

The diagnosis, differentiation, and classification of the severity of ILD and COPD rely on clinical assessment, thoracic imaging (using CT and chest radiography), and pulmonary function testing (PFT). PFT is a noninvasive method of

assessing the integrated mechanical function of the lung, chest wall and respiratory muscles. It currently comprises the gold standard for pulmonary assessment. Using PFT, the heterogeneous group of ILD typically exhibits a restrictive physiology pattern whereas COPD manifests as an obstructive physiology pattern.

PFT strictly permits a global assessment of lung functionality. In contrast, HRCT image analysis is a powerful tool with the potential for regional as well as global quantitation of pulmonary diseases. Although generally effective, radiologic interpretation of HRCT is time-consuming, largely qualitative, and prone to diagnostic variability. As a result, various CT image metrics have been proposed towards a more consistent and facilitated assessment. Early investigations into CT lung analysis employed relatively simple metrics, such as the mean attenuation value and other such first-order statistical measurements obtainable from the attenuation histogram [1–3]. More sophisticated metrics relying on texture descriptions of the parenchyma have shown promise in recent studies [4].

PFT values are used clinically to diagnose ILD and COPD. With the increasing amount of proposed image metrics, research inquiry concerns the effectiveness of these metrics compared with the gold standard PFT values. Previous research ([5–9]) investigated the correlation between various quantitative image metrics with different PFT values.

Instead of a correlation study, this paper addresses the question more from the view of feature selection. We put more emphasis on what image metrics and what PFT values can characterize ILD and COPD in a quantitative framework. Image metrics and PFT values are viewed as candidates for selecting which best characterize the corresponding diagnosis (i.e. ILD or COPD in this paper). We are interested in whether and what features from image metrics have additional information for diagnosis when PFT values are provided. This also differs from the classification work of [4] that we do not tend to train any classifiers directly. The selected features can be used as inputs for any available classifier. We use Support Vector Machine (SVM, [10]) as an example to test the efficiency of the feature selection results.

A minimal-redundancy-maximal-relevance (mRMR) information framework was introduced in [11] for such a feature selection task. The ideal selected features satisfy two constraints: maximal relevance and minimal redundancy. The *relevance* of both image and PFT features concerns the ability of such features in matching an existing classification (in our case, from clinical diagnosis). It is usually computed in terms of mutual information, correlation, or statistical tests. However, in order to get a compact subset of features to classify different types of disease, it is not enough to consider only the features with highest relevance. The selected features need to be as independent to each other as possible. This is known as the criterion of minimal *redundancy*.

In this manuscript, we provide a systematic relevancy/redundancy analysis comparing 31 various statistical image metrics and 21 PFT values obtained in patients with diagnosed ILD and COPD. The analysis framework is described in Section 2. In Section 3, we provide the results of our analysis in comparing the

characterization performance of both image and PFT features. This is followed by discussion of the results.

2 Materials and Methods

14 patients with ILD and 11 patients with COPD were retrospectively identified. Every patient underwent both thoracic HRCT image acquisition and PFT within 3 days of each other. HRCT was performed for both inspiration and expiration on a 64 multidetector row CT scanner (Siemens Medical Solutions) with reconstruction of contiguous 1 mm axial images with a B41f kernel. Inspiratory and expiratory image datasets were then analyzed through a computational software developed in our laboratory which is capable of generating several hundred distinct metrics encompassing various aspects of lung physiology (e.g. pulmonary volumetric and gross tissue indices, attenuation histogram statistics, deformation indices, co-occurrence [12] and run-length [13] matrix texture indices, and attenuation mask indices), gleaned from the relevant literature. For this study, we only focused on a portion of these metrics .

Our whole pipeline (Fig. 1) is fully automatic. First, a segmentation algorithm [14] is applied to extract lung regions and segment the trachea from CT images. Next. we compute 31 different types of image metrics (Table.2, gleaned from the relevant literature) on the expiratory and inspiratory image datasets separately. We also subtract expiratory metrics from their inspiratory counterparts to obtain additional 31 metrics. This translates into a total of 93 image metrics for each patient. We then compute the relevance of each PFT value and each image metric to the corresponding disease type. The criteria of mRMR [11] is applied to select a mixed subset of PFT values and image metrics to show the degree of redundancy between them. Finally, we train SVM classifiers using the selected feature set to test the accuracy of ILD/COPD classification.

2.1 PFT Parameters and Image Metrics

Table 1 lists all the 21 PFT values used in the study. These parameters were either directly obtained or calculated for each patient, using standard protocols following guidelines of the American Thoracic Society. As mentioned previously, 93 image metrics (shown in Table 2) were generated for each patient. We use

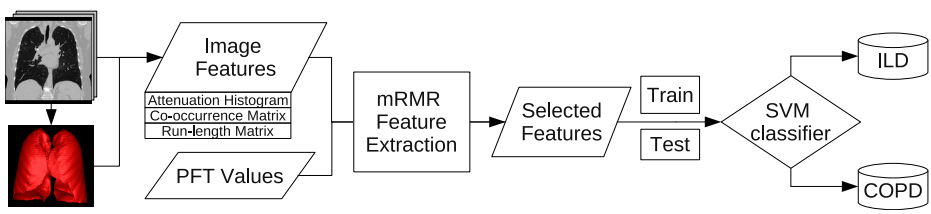


Fig. 1. Pipeline of the proposed feature selection analysis.

PFT	*FVC*	*FEV1*	*FEV1 /FVC*	*FEF max*	*FEF 25-75%*	*FEF50*	*FIF50*	*MVV*	*SVC*	*IC*	*ERV*
Rel	0.088	0.026	0.47	0.357	0.218	0.307	0.021	0.251	0.197	0.083	0.146
PFT	*TGV*	*RV (Pleth)*	*TLC (Pleth)*	*RV /TLC*	*DLCO unc*	*DLCO cor*	*DL/VA*	*VA*	*Raw*	*sGaw*	
Rel	0.268	0.268	0.35	0.218	0.041	0.15	0.066	0.269	0.101	0.074	

Table 1. All 21 PFT values (1st and 3rd rows) and corresponding relevance values (2nd and 4th rows) to the disease types.

G1 to denote the 31 features from each of the expiratory image, *G2* for the inspiratory images and *G3* for the difference of the expiratory/inspiratory metric pairs. This subset of image metrics includes those first-order statistical measures generated from the attenuation histogram. It also includes more sophisticated second-order statistical quantities related to the texture of lung parenchyma, i.e. the co-occurrence [12] and run-length [13] matrix texture indices.

2.2 Feature Selection

The primary motivation for this study is to determine how image metrics perform relative to PFT values in matching clinical diagnosis of ILD and COPD. We also select an optimal subset of mixed image metrics and PFT values under the principle of minimal redundancy and maximal relevance.

In accordance with the specifications of the algorithm in [11], each of the $N = 25$ patients is given a label c based on the clinical diagnosis, either ILD or COPD. Such a disease type c is a random variable. Also each patient has a set of $J = 93$ computed image metrics $X = \{x_1, \ldots, x_J\}$ and a set of $K = 21$ measured PFT values $Y = \{y_1, \ldots, y_K\}$. Let the mixed feature set $Z = X \cup Y$. Each element $z_i \in Z$ is also viewed as a random variable. [11] proposed to use mutual information to measure the *relevance* between z_i and c: $I(z_i, c) = \iint p(z_i, c) \log \frac{p(z_i, c)}{p(z_i)p(c)} dz_i dc$. The mRMR framework in [11] selects a subset of S from Z such that:

$$S = \arg\max_{S \subseteq Z} \left\{ \sum_{z_i \in S} I(z_i, c) - \frac{1}{\|S\|} \sum_{z_i, z_j \in S} I(z_i, z_j) \right\} \quad (1)$$

The first term in Equ. 1 maximizes the total *relevance* of the selected features with the corresponding disease types; the second term minimizes the total redundancy of all pairs of the selected features. The framework in [11] gives a heuristic way to optimize Equ.1. We use the online toolbox [18] to compute the relevance weights and select features.

Given the diagnosis c, we analyze the relevance of each feature, which indicates its individual characterization power. Then we use mRMR to select the subset of the first 25 features. If the subset includes only PFT values, this indicates that image metrics are redundant for differentiating ILD and COPD;

	No.	Metrics Type	*G1*	*G2*	*G3*
Attenuation Histogram Indices	1	lung region volume [15][5]	0.325	0.276	0.073
	2	relative volume$_{<-910}$ [5]	0.439	0.454	0.071
	3	attenuation mean [2, 3]	0.416	0.407	0.181
	4	attenuation variance [16]	0.171	0.197	0.181
	5	sum [2]	0.295	0.353	0.005
	6	attenuation skewness [4]	0.276	0.159	0.026
	7	attenuation kurtosis [4]	0.392	0.201	0.079
	8	attenuation grey level entropy [4]	0.021	0.121	0.005
	9	5% attenuation value [1]	0.47	0.527	0.353
	10	95% attenuation value [1]	0.463	0.22	0.006
	11	5% attenuation mean [1]	0.488	0.416	0.036
	12	95% attenuation mean [1]	0.055	0.065	0.071
Co-occurrence Matrix Indices	13	energy [4, 12]	0.019	0.05	0.05
	14	entropy [4, 12]	0.159	0.215	0.159
	15	correlation [4, 12]	0.003	0.001	0.002
	16	inverse difference moment [4, 12]	0.021	0.121	0.025
	17	inertia [12]	0.003	0.022	0.005
	18	cluster shade [12]	0.083	0.021	0.074
	19	cluster prominence [12]	0.113	0.074	0.104
	20	Haralick's correlation [12]	0.034	0.034	0.001
Run-length Matrix Indices	21	short run emphasis [4]	0.049	0.002	0.002
	22	long run emphasis [4]	0.074	0.002	0.002
	23	grey level non-uniformity [4]	0.243	0.316	0.007
	24	run-length non-uniformity [4]	0.325	0.353	0.006
	25	run percentage [4]	0.157	0.101	0.101
	26	low grey level run emphasis [17]	0.285	0.157	0.113
	27	high grey level run emphasis [17]	0.005	0.114	0.005
	28	short run low grey level emphasis [13]	0.285	0.157	0.058
	29	short run high grey level emphasis [13]	0.005	0.114	0.005
	30	long run low grey level emphasis [13]	0.285	0.157	0.113
	31	long run high grey level emphasis [13]	0.005	0.114	0.005

Table 2. Computed image metrics and their relevance to the disease types. *G1* denotes the metrics computed from the expiratory images; *G2* from the inspiratory images. Metrics of *G3* are generated by subtracting *G1* from *G2* . Within each of *G1/G2/G3*, the metrics are indexed from 1 to 31. The numbers in last three columns are the relevance values to the disease types of ILD and COPD.

otherwise, it indicates that image metrics have extra information that PFT values do not have for differentiating ILD and COPD.

2.3 Classification with SVMs

The last step in our analysis pipeline is to use the selected features to train classifiers to classify ILD and COPD . An efficient feature selection scheme should achieve low classification error rate with minimal number of features. We are going to study the relationship between classification error and the number of

the selected features. Also we are going to compare adding image metrics to PFT values as additional features to classification using PFT values only.

We use Support Vector Machines (SVM)[10] as classifiers in this paper. SVMs have been used widely for feature classification tasks. SVM classifies data by maximizing the margins between hyperplanes that separates data points of two classes. By choosing different kernels ([19]), data points are implicitly mapped to different higher dimensional spaces and thus better classification can be achieved. More discussion about SVMs can be found in [19].

The experiments here use SVM with Gaussian kernels to compare the selected features. We also test the linear kernel to understand whether the performance improvement is from the choice of nonlinear kernel or from the more efficient features.

3 Results

We first compare the relevance of each feature to the diagnosis of diseases. Fig. 2 graphically depicts the relevance weights of all the image metrics. The relevance weights of PFT values are plotted in Fig. 3 as a comparison. For example, the ratio of the forced expiratory volume in 1 second to the forced vital capacity (*FEV1/FVC*) has a high relevance value (0.47), as it is a meaningful clinical indication.

The relevance values from image metrics and PFT values are in a similar range. The top PFT parameter, *FEV1/FVC*, is 0.47; the top image metric, 5% attenuation value from the inspiratory images, is 0.52. This shows that some image metrics are as good indicators as PFT values for characterization of ILD and COPD. It also shows in Fig. 2 that the metrics computed from expiratory images (*G1*:the red bars) are similar to the metrics from inspiratory images (*G2*: the blue bars). Their relevance values have a high correlation of 0.84.

We use mRMR to select a subset of 250 features. As discussed before, a high relevance does not suffice for a feature to be selected. It also requires a selected feature to have low redundancy with respect to other features in the subset. In

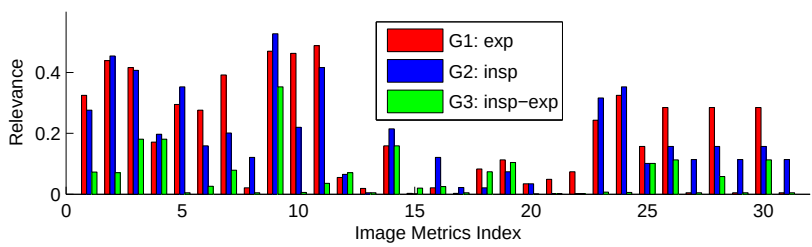


Fig. 2. Relevance (mutual information) of image metrics to different disease types. The x axis has the image metrics index listed in Table 2. Red bars are for metrics from *G1*, the expiratory images; blue bars for *G2*, the inspiratory images; and green bars for *G3*, difference of *G1* to *G2*.

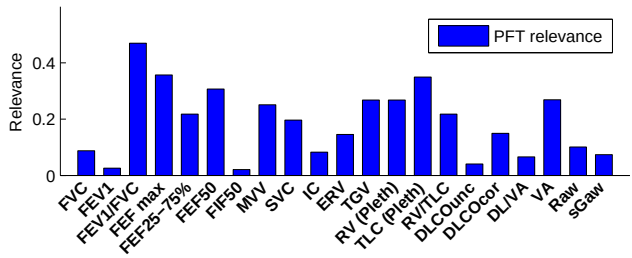


Fig. 3. Relevance (mutual information) of PFT values to different disease types. The x axis lists all the 21 PFT values in Table 1. Note that the range of the y axis is similar to the range from those image metrics in Fig. 2.

other words, a selected feature will provide additional information supporting the diagnosis that is missing from other features.

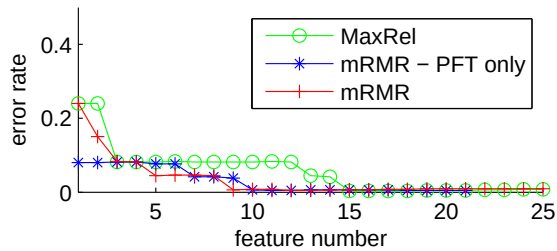
Rank	Feature	Rank	Feature
1	*G2*:5% attenuation value	14	*G3*:entropy
2	*G1*:kurtosis	15	*G1*:short run low grey level emphasis
3	*VA*	16	*G3*:sigma
4	*G1*:low grey level run emphasis	17	*G2*:mean
5	*G2*:run length nonuniformity	18	*TLC (pleth)*
6	*G1*:5% attenuation mean	19	*G2*:short run low grey level emphasis
7	*FEF max*	20	*G2*:sum
8	*G1*:95% attenuation value	21	*G2*:5% attenuation mean
9	*G1*:cluster prominence	22	*G3*:relative volume$_{<-910}$
10	*G1*:emphysema relative ratio	23	*G2*:relative volume$_{<-910}$
11	*G1*:volume	24	*G3*:cluster prominence
12	*G2*:grey level nonuniformity	25	*G1*:long run low grey level emphasis
13	*FEV1/FVC*		

Table 3. The first 25 selected features from mRMR [11]. Note these features include both PFT values and image metrics. The image metrics includes both the first order statistical measurements obtainable from the attenuation histogram and other more sophisticated metrics of texture descriptions.

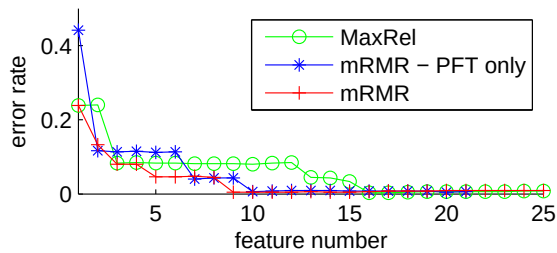
Table 3 lists all the 25 selected features using mRMR. Both PFT values and image metrics are selected. This indicates that image metrics provide extra information for differentiation of ILD and COPD, comparing with using only PFT values. Also image features computed both from the first-order statistical measurements (like the attenuation value) and from the texture descriptions (like the grey level run emphasis and the cluster prominence) are selected in the same subset. This also suggests that both types of image features are valuable for diagnosis. Out of these 25 features, 20 features statistically differ in ILD from

COPD groups using an unpaired t-test (null hypothesis rejected at 2% level), which means that statistical significance exists on our relatively small datasets.

In comparison to feature selection using mRMR, the baseline used in the paper is to rank features by their relevance values (i.e. $I(z_i, c)$ in Equ.1) and such a criterion is denoted as MaxRel as in [11]. In this paper, we use SVMs as classifiers to compare the efficiency of the first $n = 1$ to 25 features ranked by mRMR and MaxRel by increasing the size (n) of the selected feature set. In each test, one subject is excluded when training SVMs and that subject is used as the test set. The error rates are computed as the average ratio of wrongly classified subject by repeating using each of the patients as the test subject.



(a) SVM with Gaussian kernel



(b) SVM with linear kernel

Fig. 4. Classifying ILD and COPD using the first $n = 1$ to 25 selected features with (a) SVM with Guassian kernel and (b) SVM with linear kernel, using MaxRel, mRMR and mRMR with only PFT values. For each curve, the x axis is the number of features; the y axis is the error rate. One subject is excluded when learning SVM parameters and that subject is used to test the accuracy of the learned SVM.

We test two SVM models with different kernels. The first is SVM with Gaussian kernel, which is a popular choice for general classification tasks [19]. Features are normalized by dividing them by the maximum the absolute values. To further investigate the efficiency of the selected features and isolate the influence of choice of kernels, we also test SVM with the most simple linear kernel. As shown in both Fig. 4 (a) and (b), mRMR (red curves with crosses) reaches low error rate using 9 features in comparison to MaxRel using 15 (SVM with Gaus-

sian kernel) or 16 (SVM with linear kernel). This shows the efficiency of mRMR framework.

To further demonstrate that image features are not redundant to the PFT values, we apply mRMR only on the subset of 21 PFT values and have a similar experiment by computing error rates with increasing selected feature set size. The performance is shown as "mRMR - PFT only" (the blue curves with stars) in Fig. 4. Although PFT values have similar error curves using SVM with Gaussian kernels (Fig. 4 (a), it is inferior using linear kernels (Fig. 4). This can be explained as the selected image features are able to compensate the inefficiency of PFT features.

4 Conclusions

Various image metrics have been proposed and used in the literature to differentiate ILD and COPD. In this paper we studied their relevance values to the corresponding diagnoses in comparison with PFT values. Using a minimal-redundancy-maximal-relevance (mRMR) framework, we further looked into the redundancy between image metrics and PFT values. The result shows that not only some image metrics have similar relevance values compared with PFT, but also image metrics are not redundant when PFT values are provided. Image metrics of the first order statistics from the attenuation histogram and of more sophisticated texture descriptions are both selected, which again suggests that these two types of image metrics are both valuable for a further investigation. We finally gave a selected optimal subset of 25 features for characterization of ILD and COPD, including both image metrics and PFT values.

It should be noted that we did not go far enough in achieving an automoated diagnosis of ILD and COPD system in this paper. Although certainly plausible in the future, such an automated diagnosis system, however, is premature at this point and not what we intended to develop in this paper. Rather, the focus of this paper is the demonstration that various image metrics are capable of providing significant information in characterizing clinical diagnosis compared to the gold standard of PFT values for ILD and COPD pathologies.

References

1. Mller, N.L., Staples, C.A., Miller, R.R., Abboud, R.T.: "density mask". an objective method to quantitate emphysema using computed tomography. Chest **94**(4) (Oct 1988) 782–787
2. Gevenois, P.A., Vuyst, P.D., Sy, M., Scillia, P., Chaminade, L., de Maertelaer, V., Zanen, J., Yernault, J.C.: Pulmonary emphysema: quantitative CT during expiration. Radiology **199**(3) (Jun 1996) 825–829
3. Lamers, R.J., Kemerink, G.J., Drent, M., van Engelshoven, J.M.: Reproducibility of spirometrically controlled CT lung densitometry in a clinical setting. Eur Respir J **11**(4) (Apr 1998) 942–945

4. Xu, Y., Sonka, M., McLennan, G., Guo, J., Hoffman, E.A.: MDCT-based 3-D texture classification of emphysema and early smoking related lung pathologies. IEEE Trans Med Imaging **25**(4) (Apr 2006) 464–475
5. Zaporozhan, J., Ley, S., Eberhardt, R., Weinheimer, O., Iliyushenko, S., Herth, F., Kauczor, H.U.: Paired inspiratory/expiratory volumetric thin-slice CT scan for emphysema analysis: comparison of different quantitative evaluations and pulmonary function test. Chest **128**(5) (Nov 2005) 3212–3220
6. Coxson, H.O., Rogers, R.M., Whittall, K.P., D'yachkova, Y., Par, P.D., Sciurba, F.C., Hogg, J.C.: A quantification of the lung surface area in emphysema using computed tomography. Am J Respir Crit Care Med **159**(3) (Mar 1999) 851–856
7. Stavngaard, T., Shaker, S.B., Bach, K.S., Stoel, B.C., Dirksen, A.: Quantitative assessment of regional emphysema distribution in patients with chronic obstructive pulmonary disease (copd). Acta Radiol **47**(9) (Nov 2006) 914–921
8. Madani, A., Zanen, J., de Maertelaer, V., Gevenois, P.A.: Pulmonary emphysema: objective quantification at multi-detector row CT–comparison with macroscopic and microscopic morphometry. Radiology **238**(3) (Mar 2006) 1036–1043
9. Temizoz, O., Etlik, O., Sakarya, M.E., Uzun, K., Arslan, H., Harman, M., Demir, M.K.: Detection and quantification of the parenchymal abnormalities in emphysema using pulmo-CT. Comput Med Imaging Graph **31**(7) (Oct 2007) 542–548
10. Vapnik, V.N.: The nature of statistical learning theory. Springer-Verlag New York, Inc. (1995)
11. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell **27**(8) (August 2005) 1226–1238
12. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics. **3**(6) (1973) 610–621
13. Dasarathy, B.R., Holder, E.B.: Image characterizations based on joint gray-level run-length distributions. Pattern Recognit. Lett. **12** (1991) 497–502
14. Hu, S., Hoffman, E.A., Reinhardt, J.M.: Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. IEEE Trans Med Imaging **20**(6) (Jun 2001) 490–498
15. Hoffman, E.A., Simon, B.A., McLennan, G.: State of the art. a structural and functional assessment of the lung via multidetector-row computed tomography: phenotyping chronic obstructive pulmonary disease. Proc Am Thorac Soc **3**(6) (Aug 2006) 519–532
16. Stolk, J., Putter, H., Bakker, E.M., Shaker, S.B., Parr, D.G., Piitulainen, E., Russi, E.W., Grebski, E., Dirksen, A., Stockley, R.A., Reiber, J.H.C., Stoel, B.C.: Progression parameters for emphysema: a clinical investigation. Respir Med **101**(9) (Sep 2007) 1924–1930
17. Chu, A., Sehgal, C.M., Greenleaf, J.F.: Use of gray value distribution of run lengths for texture analysis. Pattern Recognition Letters **11** (1990) 415–420
18. http://research.janelia.org/peng/proj/mRMR/
19. Schölkopf, B., Smola, A.J.: Learning with kernels : support vector machines, regularization, optimization, and beyond. Adaptive computation and machine learning. MIT Press (2002)

Graph Cut-based Automatic Segmentation of Lung Nodules using Shape, Intensity, and Spatial Features

Xujiong Ye, Gareth Beddoe and Greg Slabaugh

Medicsight PLC, London, UK

Abstract. This paper presents a new, fully automatic method of accurately extracting lesions from CT data. It first determines, at each voxel, a five-dimensional feature vector that contains intensity, shape index, and 3D spatial location. Then, non-parametric mean shift clustering is applied to produce intensity and shape mode maps. Finally, a graph cut algorithm segments the image using a novel energy formulation that incorporates shape, intensity, and spatial features. A key difference from the usual graph construction is that we connect modes (small regions, or super-pixels resulting from mean shift clustering) instead of pixels. The initial foreground and background can be automatically obtained by calculating highly spherical regions based on the shape index map. The proposed method has been evaluated on a clinical dataset of thoracic CT scans that contains 100 nodules. A volume overlap ratio between each segmented nodule and the ground truth annotation is calculated. Using the proposed method, the mean overlap ratio over all the nodules is 0.81. On visual inspection as well as using a quantitative evaluation, the experimental results demonstrate the potential of the proposed method. The rich information provided by the joint spatial–intensity-shape features provides a powerful cue for successful segmentation of nodules adjacent to structures of similar intensity but different shape.

1 Introduction

Accurate and automatic segmentation of medical images is an essential component of a computer-aided diagnosis (CADx) system. However, medical image segmentation is typically a difficult task due to noise resulting from the image acquisition process, as well as the characteristics of the object itself and its neighborhood. Lesions may be embedded in areas of complicated anatomy; and may have very similar intensities to their adjacent tissues (e.g. juxta-vascular nodules). In such cases, traditional intensity-based or model-based methods may fail to properly segment the object [1-3]. For example, a contrast-based region growing approach was introduced in [1]. This method assumes that the region of interest appears as a bright or dark object relative to the surrounding tissue. However, since adjacent blood vessels have a similar intensity to the nodule, the segmentation tends to include a part of a blood vessel along with the nodule. A morphological approach was presented in [2]. One issue with this method is its sensitivity to the morphology template size, which makes it difficult to choose a suitable template for all different kinds of nodules. As indicated by the authors, this algorithm is targeted for small and high contrast nodules; thus for a large nodules, especially when a blood vessel is attached to a nodule, the algorithm might fail to properly delineate the nodule.

Image segmentation methods based on energy minimization have been intensively researched [4-10]. In particular, graph cut based methods, which can achieve a global minimum of energy functions used in image segmentation, have been shown much promise in medical image computing. However, in most graph-based methods, the graph vertices are constructed at image pixels, and the segmentation energy is composed of intensity terms. For example, Zheng *et al.* [5] proposed a framework to simultaneously segment and register the lungs and nodules in CT data. For segmentation, a 2D pixel based graph cut algorithm was applied on the 3D lung and nodule datasets. It is noted that, by representing a graph vertex using an image pixel, the number of nodes in the graph increases polynomically with the image size (N^D, where N is the number pixels in one of the D dimensions); this dramatically increases the computation time. To improve efficiency, Li *et al.* [6] introduced a graph built on a pre-segmentated image using a watershed algorithm. However, their graph cut formulation is solely based on the image intensities. It is known that pixel intensity can be locally erroneous due to noise and other image acquisition issues (such as Partial Volume Effect (PVE) in CT). Thus, in these cases, noise can adversely affect the performance of a graph-based segmentation. Slabaugh *et al*. [7] incorporated an elliptical shape prior into the graph-cut segmentation framework. Xu *et al*. [8] presented a graph-cuts based active contours approach to object segmentation method. Zheng *et al*. [9] constructed a graph Laplacian matrix for the estimation of Ground-Glass Opacity (GGO) nodule in CT. Recently, Liu *et al*. [10] applied ordering constraints into an energy smoothness term based on an initial labeling. A simple geometric shape prior was also incorporated in a graph cut segmentation.

In this paper, we propose an automatic mode-based graph cut method for lung nodule segmentation. An overview of the approach appears in Fig. 1. At each voxel in the image, the volumetric shape index (SI) is computed. The shape index, along with the image intensity and spatial position (x, y, z) are concatenated into a five dimensional feature vector at each voxel. In this five dimensional joint spatial-intensity-shape (JSIS) feature space, mean shift clustering is applied, producing intensity and shape index mode maps (super-pixels). Then, a graph is constructed using the super-pixels as vertices. Weights in the graph are computed using a novel energy formulation that considers not only image intensity but also the shape feature. The use of mean-shift generated super-pixels produces better results and improves speed significantly compared to a dense graph with vertices at every voxel. The experimental results on CT lung nodules demonstrate the high performance of the proposed method.

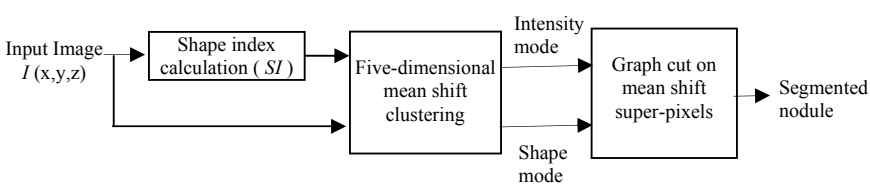


Fig. 1. Flow diagram of the proposed graph cut based method

2 Mean Shift Clustering of JSIS features

Our approach first computes the JSIS features, which are then clustered in a five-dimensional space using mean shift. In this section, we review the shape index feature and our mean shift approach.

2.1 Volumetric shape index: a 3D geometric feature

The volumetric shape index (SI) at voxel $p(x, y, z)$ can be defined as [11][12]:

$$SI(p) = \frac{1}{2} - \frac{1}{\pi} \arctan \frac{k_1(p) + k_2(p)}{k_1(p) - k_2(p)} \quad (1)$$

where $k_1(p)$ and $k_2(p)$ are the principal curvatures at voxel p, which are defined as:

$$k_1(p) = H(p) + \sqrt{H^2(p) - K(p)}, \quad k_2(p) = H(p) - \sqrt{H^2(p) - K(p)}$$

where $K(p)$ and $H(p)$ are the Gaussian and mean curvatures, respectively.

The calculation of the Gaussian and mean curvatures are based on the first and second fundamental forms of differential geometry. A practical approach for computing these forms is to use the smoothed first and second partial derivatives the image with respect to x, y, z as suggested in [13]. In this paper, prior to shape index calculation, a single-scale Gaussian smoothing is employed to obtain the smoothed image with standard deviation of 1.5.

Shape index represents the local shape feature at each voxel. Every distinct shape, except for the plane, corresponds to a unique shape index. For example, the shape index value of 1.00 indicates a sphere-like shape (e.g. nodule), and 0.75 indicates a cylinder-like shape (e.g. vessel). Based on the definition, volumetric shape index directly characterizes the topological shape of an iso-surface in the vicinity of each voxel without explicitly calculating the iso-surface. This feature provides rich information for automated segmentation of anatomical structures or lesions in medical images, where the region of interest is within an area of complicated anatomy and image intensities of different shapes are similar to each other (such as an adjoining lung nodule).

2.2 Combination of shape index feature into mean shift framework

For each voxel, 3D spatial location, intensity and volumetric shape index features are concatenated in the joint spatial-intensity-shape (JSIS) domain of dimension d=5. Given n data points p_i, i=1,…,n in a 5-dimensional space, the multivariate kernel is defined as the product of three radially symmetric kernels:

$$K_{h_s h_r h_{si}}(f) = c_{k,5} k\left(\left\|\frac{f^s}{h_s}\right\|^2\right) \cdot k\left(\left\|\frac{f^r}{h_r}\right\|^2\right) \cdot k\left(\left\|\frac{f^{si}}{h_{si}}\right\|^2\right) \quad (2)$$

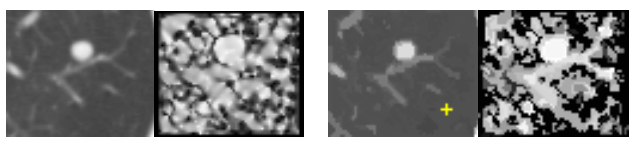
where $c_{k,5}$ is a normalization constant which assures *K(f)* integrates to 1. f^s is the spatial location, f^r is the intensity and f^{si} is the shape index feature; h_s, h_r and h_{si} are the kernel window size for spatial, intensity and shape index kernel function. The normal kernel is used in this paper, where $k(f) = (2\pi)^{-d/2} \exp\left(-1/2\|f\|^2\right)$.

By using the mean shift framework [14], the shape index feature can be combined with the intensity for clustering. The mean shift vector with three kernel windows (spatial, intensity and shape index) can then be calculated as:

$$m_{h_s h_r h_{si}}(p_i) = \frac{\sum_{i=1}^{n} x_i g\left(\left\|\frac{f^s}{h_s}\right\|^2\right) \cdot g\left(\left\|\frac{f^r}{h_r}\right\|^2\right) \cdot g\left(\left\|\frac{f^{si}}{h_{si}}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{f^s}{h_s}\right\|^2\right) \cdot g\left(\left\|\frac{f^r}{h_r}\right\|^2\right) \cdot g\left(\left\|\frac{f^{si}}{h_{si}}\right\|^2\right)} - p_i \quad (3)$$

where $g(f) = -k'(f)$. The mean shift vector always points toward the direction of the maximum increase in the density function.

It is noted that the mean shift algorithm estimates the modes (the densest regions) of the multivariate distribution underlying the feature space. The set of points that converge to the same mode is defined as the attraction basin. Mean shift maps all the data samples to the local maxima of their corresponding attraction basin based on five-dimensional features. Super-pixels (modes) are formed for the set of pixels in each attraction basin. Each super-pixel has a constant shape index and intensity. This produces mode maps, namely an intensity mode map (M_I), and a shape index mode map (M_{SI}), and a spatial mode map (M_S). The spatial mode map is not used directly in our energy function; however, spatial information is utilized when defining neighbors in our graph. Fig. 2 shows a nodule attached to a vessel and its corresponding intensity and shape index mode maps. It can be noted that the mode maps ((c) and (d)) from JSIS mean shift clustering can be seen as "filtered" images and are less contaminated by outliers. In the following section, graph cut based segmentation is applied on these super-pixels.



(a) (b) (c) (d)

Fig.2. One attached nodule with its intensity and shape mode maps (a) Original CT sub-image; (b) Shape index map based on Eq. (1); (c) Intensity mode and (d) shape index mode maps.

3 Automatic Graph Cut based Segmentation on Mean Shift Mode Map with Shape Feature

In this section, we consider the mode map as a graph *G*. As mentioned above, the graph $G=(V, E)$ is defined with vertices $v \in V$ representing super-pixels determined

from five dimensional mean-shift clustering, and edges $\varepsilon \in E$ connecting adjacent super-pixels. A key difference from the usual graph construction is that we connect super-pixels instead of the original pixels. As a result, the number of vertices in G is greatly reduced compared to the original number of pixels in the image. Two key issues are addressed in the following two sub-sections: initialization and energy function definition.

3.1 Initialization based on high spherical concentration

In our previous work, we have developed an automatic lung nodule detection algorithm [15], which produced a small number of potential nodule regions. The aim of this paper is to accurately delineate each nodule boundary. Given that a nodule is generally either spherical or has local spherical elements, while a blood vessel is usually oblong, for each potential nodule region (or region of interest), a spherical concentration is calculated for the automatic estimation of initial nodule (foreground) region.

A cluster of high shape index voxels is automatically determined using hysteresis thresholding [13] applied to the shape index map. This algorithm finds a spherical region $\Re_s$ formed of 3D connected voxels that all have shape index greater than or equal to a relaxed threshold and contain voxels with a shape index greater than or equal to a high threshold. The voxels in this spherical region define the foreground seeds for the graph cut segmentation.

To produce the background seeds, the foreground region is enlarged based on the distance transform [16]. The initial background region can be obtained by inverting the enlarged foreground region. Here, the foreground region is enlarged to ensure that the background seeds do not cover the nodule to be segmented. Fig.3 shows an example of the segmented vascular nodule using the above automatic calculation of foreground and background seeds, where, the high threshold and relaxed threshold for hysteresis thresholding were chosen to be 0.92 and 0.82 respectively; and the initial foreground (Fig.4 (a2 and b2)) was enlarged 10 layers based on the distance transform to obtain the background (Fig.4(a3 and b3)).

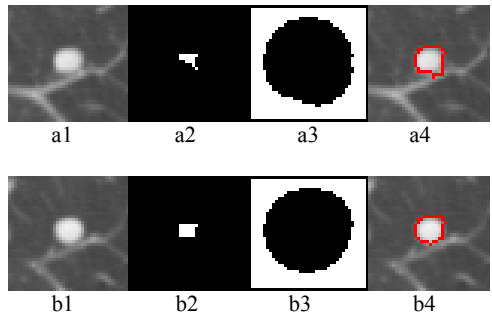


Fig.3. An example of one attached solid nodule segmentation based on the automatic calculation of initial foreground and background. (a1-b1) 3D nodule in 2 continuous slices in CT; (a2-b2) initial foreground based on high spherical concentration; (a3-b3) initial background; (a4-b4) segmentation results by the proposed graph cut based method.

3.2 Energy function

The lesion segmentation problem is formulated as a binary labeling problem, so the goal is to assign a unique label $l_i \in \{0,1\}$ to each super-pixel (mode) (where 0 is background and 1 stands for foreground (lesion)) by minimizing a Gibbs energy *E(L)* [17]:

$$E(L) = \sum_{i \in V} E_1(l_i) + \lambda \sum_{(i,j) \in \varepsilon} E_2(l_i, l_j) \quad \textbf{(4)}$$

where $E_1(l_i)$ is the data energy, determining the energy to assign the label l_i to the mode *i*, and $E_2(l_i, l_j)$ is the smoothing energy, denoting the cost of assigning the labels l_i and l_j to adjacent super-pixels *i* and *j*. λ is a weighting factor. The details of energy minimization via the graph cut algorithm for binary labeling can be found in [1]. Below we focus on how to define the two energy terms.

Data energy (t-link energy): Given the initial foreground $\{M_m^F\}$ and background regions $\{M_n^B\}$ which are automatically calculated based on the hysteresis thresholding of shape index map discussed in the section 3.1. Here, *m* and *n* are the super-pixel indices for initial foreground and background, respectively. For each super-pixel *i*, the intensity distance of the super-pixel to the foreground super-pixels $\{M_m^F\}$, weighted by the shape feature, is calculated as:

$$d_i^F = \min_m \left\| I_{\mathrm{mode}}(i) - M_m^F \right\| \times \left(1 - \exp\left(-\frac{(SI_{\mathrm{mode}}(i) - 1.0)^2}{\sigma_a^2} \right) \right),$$

where $\sigma_a = 0.3$, $I_{\mathrm{mode}}(i)$ and $SI_{\mathrm{mode}}(i)$ are the i*th* super-pixel's intensity and shape index resulting from mean shift clustering. Alternatively, an exponential function can

also be used for the calculation of the intensity similarity. The intensity distance to the background super-pixels $\{M_n^B\}$ is as: $d_i^B = \min_n \|I_{\text{mode}}(i) - M_n^B\|$. Therefore, $E_1(l_i)$ is defined as:

$$E_1(l_i = 1) = \frac{d_i^F}{d_i^F + d_i^B} \qquad E_1(l_i = 0) = \frac{d_i^B}{d_i^F + d_i^B} \tag{5}$$

We note that a nodule is generally either spherical or has local spherical elements (we define *spherical elements* as a local grouping of voxels recognized by high volumetric shape index values). It can be seen that d_i^F in Equation (5) encourages a super-pixel to have the same label as the initial foreground super-pixels if it has a similar intensity to the foreground super-pixels and also a shape feature closer to one.

Smooth energy with shape feature (l-link): The second term $E_2(l_i, l_j)$ in Equation (4) is defined as:

$$E_2(l_i, l_j) = w_{si}(E_I + E_{SI}) + (1 - w_{si}) \cdot E_I \cdot E_{SI} \tag{6}$$

where E_I is the intensity energy term, denoting the intensity difference between two adjacent super-pixels *i* and *j*, which is defined as: $E_I(l_i, l_j) = 1/(\|I_{\text{mode}}(i) - I_{\text{mode}}(j)\| + 1)$. This means super-pixels with similar intensities have a larger E_I, which leads to assigning the same labels to the two super-pixels.

E_{SI} is the shape energy term, denoting the shape difference between two adjacent super-pixels, which is defined as $E_{SI}(l_i, l_j) = 1/(\|SI_{\text{mode}}(i) - SI_{\text{mode}}(j)\| + 1)$. Similar to the intensity term, the shape energy term captures the shape features for the two adjacent super-pixels, if both the shape index values are similar, E_{SI} is larger, which means with high probability, both super-pixels have the same label. It is noted that, both the intensity mode value and shape mode value have been normalized to the same scale for the calculation of intensity energy term and shape energy term.

It can be seen from Equation (6), the intensity term and shape term are combined through a weighting function w_{si} which is defined as:

$$w_{si}(SI_{\text{mode}}(i), SI_{\text{mode}}(j)) = \exp\left(-\left(\frac{SI_{\text{mode}}(i) - SI_{\text{mode}}(j)}{\sigma_{si}}\right)^2\right), \text{ here } \sigma_{si} = 0.3$$

It is noted that, when the shapes between two adjacent super-pixels are the same, $w_{si} = 1$, and the energy depends on the first term of Equation (6). However, when two adjacent super-pixels have very different shapes, w_{si} is small, so the Equation 6 depends on the second term. Fig.4 shows a juxta-vascular nodule segmentation by using different smooth energy functions. Due to PVE in CT imaging, part of the nodule's pixels (e.g. Fig.4(a3)) have relatively low intensities, compared to that on the other slices. By using the first term only in Equation (6), those pixels with low intensity (but similar shape feature) can still be correctly identified as being part of the nodule object as seen in Fig.4 (b3). However, some small amounts of vessel (similar intensity but different shape feature) are also included into the nodule object, as seen in Fig. 4(b1-b3). This is because the first term equally considers the similarity for both of the intensity and shape features. Fig.4(c1-c3) are the results by using the second term only. It can be seen that, the shape feature is only used as a weighting to

the intensity feature. Different shapes gives lower weighting to intensity term, this is why part of nodule can be properly separately from the adjoining vessel with similar intensity, as shown in Fig.4(c1-c2). However, the pixels with lower intensity due to PVE are wrongly identified as background due to the different intensities, compared to that on the other slices, as shown in Fig.4(c3). Fig. 4(d1-d3) are the results by combining both of terms as in Equation (6), in which the nodule boundary can be properly delineated despite the PVE and the presence of vessels with similar intensity.

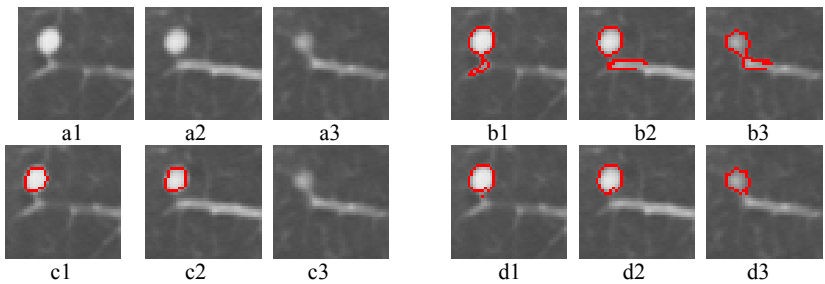


Fig.4. An example of one attached solid nodule segmentation by using different smooth energy function. (a1-a3) 3D nodule in 3 continuous slices in CT; (b1-b3) nodule segmentation by using the first term only in Equation 6; (c1-c3) results by using the second term only in Equation 6; (d1-d3) nodule segmentation by using the smooth energy in Equation 6.

4 Experimental Results and Discussion

The proposed algorithm has been evaluated on CT lung data. The three kernel window sizes (spatial h_s, intensity h_r and shape index h_{si}) in the five-dimensional mean shift clustering were set to be 3.0, 6.5 and 3.0, respectively. The proposed graph cut algorithm was applied to the mean-shift super-pixels using Equations 4, 5, and 6, where the weighting factor λ was set to be 100.

Fig.5 shows an example of the proposed method on one Ground-Glass Opacity (GGO) nodule. It is known that GGO nodules are usually with faint contrast, irregular shape and fuzzy margins, it is challenging to properly segment the nodule boundary. The proposed method demonstrates good performance on this GGO nodule segmentation. For comparison, the segmentation results without the shape feature are given in Fig. 5(a3) and (b3), where, four-dimensional mean shift with spatial and intensity features were used, also in the definition of smooth energy term (6), only the intensity energy term was considered. It can be seen that, by considering the shape index feature in both the mean shift clustering and the definition of energy formulation for graph cut, the nodule boundary can be properly delineated from the background despite the presence of other non-target structures (such as vessels).

The performance of our mode-based (mode map) graph cut algorithm was also compared with that of pixel-based method, where, in the graph construction, each vertex represents one pixel. Fig.6 shows the comparison results on another GGO nodule image, where, 46195 vertices (pixels) was constructed in the pixel-based graph, compared to 946 vertices for the super-pixel based method. Testing was performed on a system with a 2.39GHz CPU and 2GB memory. Construction of the

graph and energy minimization required 16 seconds for the pixel-based method, and 1.2 seconds for the super-pixel based method (including the mean shift). The majority of the computation time is in the graph construction, which includes the calculation of the both energy terms for each vertex. On the super-pixel based graph, the fewer vertices results in a much faster run-time.

Since the intensity mode map from the five-dimensional joint spatial-intensity-shape index mean shift algorithm expresses the local structure of the data, it can be seen that our proposed super-pixel-based method produces better results and improves the speed significantly.

For a quantitative evaluation, the proposed method has been tested on a database of clinical chest CT scans, containing 100 nodules (solid and mixed-sold nodules) with a slice thickness ranging from 0.5mm to 2.0mm. The size of the nodules ranged between 5mm to 20mm in diameter. To produce the ground truth, each nodule boundary was manually delineated by experienced radiologists. An overlap ratio between the segmented nodule and the ground truth annotation is calculated. Fig. 7 shows the overlap ratios based on the proposed method with and without shape index feature. It is noted that, without shape features, the mean overlap ratio for the whole dataset is 74% with standard deviation (*std*) of 0.08. However, the mean overlap ratio has been increased to 81% with the *std* decreasing to 0.047 by using the proposed method. This indicates the segmentation based on our proposed method is stable and accurate for different types nodules.

In this paper, the parameters for three kernel window size (spatial h_s, intensity h_r, and shape index h_{si}) in Equation 2 and the weighting factor (λ) in Equation 4 are chosen experimentally. As it is noted that, kernel window sizes depend on the data structure. To improve the performance, we are currently analyzing the sensitivity of the segmentation results to those parameters and also a variable window size needs to be further investigated.

Generally speaking, it is challenging to segment GGO boundary due to its irregular shape and faint contrast. In this section, Fig.5 and Fig.6 show good examples of GGO nodule segmentation by using the proposed method. This is because, firstly, the method clusters pixels under mean shift framework by taking into account the joint spatial-intensity-shape feature. The resulting mode map significantly reduces the variance of both of the intensity and shape features. Then, the super-pixels from mean shift clustering are further merged by using graph cut algorithm. However, this is a pilot study for the GGO nodule segmentation. Further experiments (e.g. quantitative evaluation) are needed before the method can be applied in clinical practice.

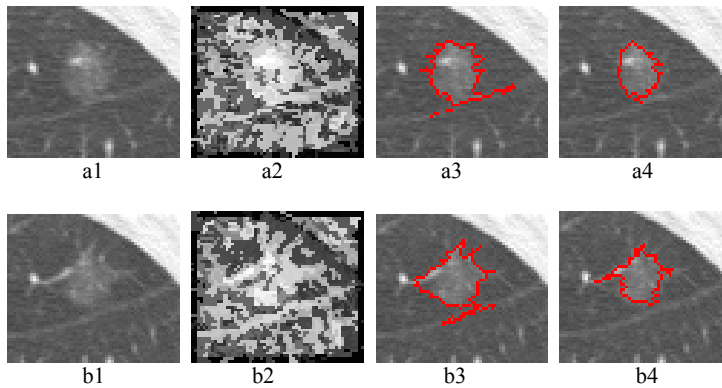


Fig.5. Example of one GGO nodule segmentation. (a1-b1): 3D GGO in two continuous slices; (a2-b2): shape index mode map from five-dimensional mean shift; (a3-b3): segmentation without shape feature; (a4-b4): segmentation results based on the proposed method.

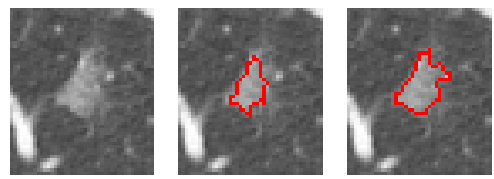


Fig.6. Segmentation result on one GGO (left) based on pixel-based graph-cut algorithm (middle) and our proposed mode-based graph-cut algorithm (right).

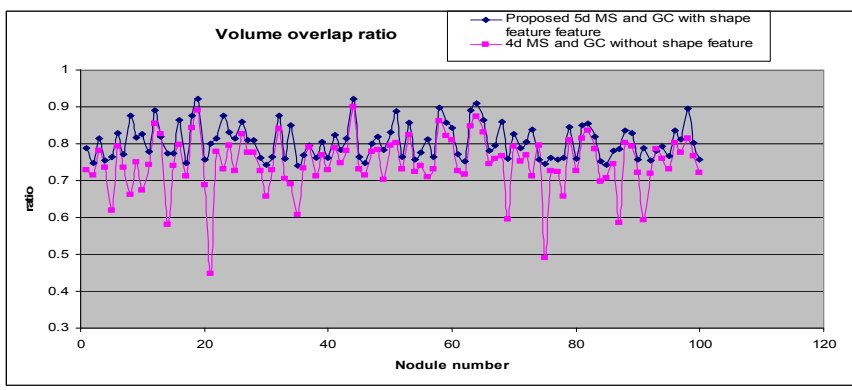


Fig.7. Volume overlap ratio based on the two different methods

5 Conclusion

We have presented a new automatic method of extracting lung nodules from CT data. A five dimensional JSIS mean shift clustering is firstly used to produce both of intensity and shape index mode maps. A graph cut algorithm is then applied to the mode map using a novel energy formulation which considers not only image intensity but also the shape feature. The joint JSIS feature provides rich information for lesion segmentation. Both by visual inspection on both solid nodules (such as Fig.3 and 4) and GGO nodules (such as Fig.5 and 6), as well as using a quantitative evaluation on 100 nodules (solid and mixed-solid) demonstrates the potential of the proposed method. The method can not only successfully segment nodules adjacent to structures of similar intensity but different shape, but also can correctly identify some part of nodules with different intensity (due to PVE in CT imaging) but similar shape.

References

1. S.A.Hijjatoleslami and J.Kitter, "Region growing: A new approach," IEEE Trans. Image Process., vol.7, no.7, pp.1079-1084, (1998).
2. W.J.Kostis, A.P.Reeves, D.F.Yankelevitz, and C.I.Henschke, "Three–dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images", IEEE Trans. Med. Imag., vol.22, no.10, pp.1259-1274, (2003).
3. W.Mullally, M. Betke, J.Wang, and J.P.Ko, "Segmentation of nodules on chest computed tomography for growth assessment," Med. Phys., vol.31, no.4, pp.839-848, (2004).
4. Y.Boykov, O.Veksler, R.Zabin, "Fast Approximate Energy Minimization via Graph Cuts", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.23, No.11, (2001).
5. Y.Zheng, K.Steiner, T.Bauer, J.Yu, D.Shen, "Lung nodule Growth Analysis from 3D CT Data with a Coupled Segmentation and Registration Framework", ICCV, (2007).
6. Y. Li, J.Sun, C.-K.Tang and H.-Y.Shum. "Lazy snapping", ACM Trans. Graphics 23(3):303-308, (2004).
7. G.G.Slabaugh and G.B.Unal. Graph cuts segmentation using an elliptical shape prior. In ICIP(2), pp.1222-1225, (2005).
8. N.Xu, N.Ahuja, R.Bansal, "Object segmentation using graph cuts based active contours", Computer Vision and Image Understanding, 107, pp.210-224, (2007).
9. Y.Zheng, C.Kambhamettu, T.Bauer and K.Steiner, "Estimation of Ground-Glass Opacity Measurement in CT Lung Images", MICCAI, pp.238-245, (2008).
10. X.Liu, O.Veksler and J.Samarabandu, "Graph Cut with Ordering Constraints on Labels and its Applications", CPVR, Alaska, (2008).
11. O.Monga and S. Benayoun, "Using partial derivatives of 3D images to extract typical surface features," Computer Vision and Image Understanding, vol.61, pp.171-189, (1995).
12. O.Faugeras, "Three-dimensional computer vision: A geometric view-point", Cambridge, MA: MIT press, (1993).
13. H.Yoshida and J. Nappi, "Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps," IEEE Trans. Medical Imaging, 20(12), pp.1261-1273, (2001).
14. D.Comaniciu and P.Meer, "Mean shift: A robust approach toward feature space analysis," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.24, pp.603-619, (2002).
15. X.Ye, X.Lin, J.Dehmeshki, G.Slabaugh, G.Beddoe, "Shape-based computer-aided detection of lung nodules in thoracic CT images," IEEE Trans. Bio-medical engineering, 56(7), 1810-20, (2009).
16. R.C, Gonzalez and R.E. Woods, Digital Image Processing, Addison Wesley, (2002).
17. S.Geman and D.Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images." Trans. Pattern Analysis and Machine Intelligence, 6:721-741 (1984).

Automated Quantitative Analysis of a Mouse Model of Chronic Pulmonary Inflammation using Micro X-ray Computed Tomography

Xabier Artaechevarria[1], Daniel Pérez-Martín[1], Joseph M. Reinhardt[2], Arrate Muñoz-Barrutia[1] and Carlos Ortiz-de-Solórzano[1]

[1] Center for Applied Medical Research, University of Navarra, Pamplona, Spain
[2] Biomedical Engineering Department, University of Iowa, Iowa City, Iowa, USA

Abstract. Micro-CT has emerged as an excellent tool for in-vivo imaging of the lungs of small laboratory animals. Several studies have shown that it can be used to assess the evolution of pulmonary lung diseases in longitudinal studies. However, most of them rely on non-automatic tools for image analysis, or are merely qualitative. In this article, we present a longitudinal, quantitative study of a mouse model of silica-induced pulmonary inflammation. To automatically assess disease progression, we have devised and validated a lung segmentation method that combines threshold-based segmentation, atlas-based segmentation and level sets. Our volume measurements, based on the automatic segmentations, point at a compensation mechanism which leads to an increase of the healthy lung volume in response to the loss of functional tissue caused by inflammation.

1 Introduction

Lung cancer and chronic obstructive pulmonary disease (COPD) have high prevalence and mortality in the developed world. Recent studies point at possible common causative mechanisms between these two diseases, related to the inflammatory process triggered by external factors, mainly cigarette smoke [1, 2].

Therefore, novel experimental models are being used to study the role of inflammation in lung cancer and COPD. Among them, in vivo experiments using small animals are likely to play an important role, since they represent an essential intermediate step between the in vitro experiments and the studies on human subjects.

High resolution X-ray computed tomography (micro-CT) can be used to non-invasively discriminate tissue composition in small animals, based on the different X-ray absorption coefficients of the tissues. It is thus very well suited for imaging the lung, because the air that fills this organ provides high contrast with the surrounding tissues [3]. Giving these facts, micro-CT can be efficiently used in longitudinal experiments aimed at studying the evolution of pulmonary diseases. Namely, several previous works have shown the usefulness of micro-CT

for the study of small animal models of lung diseases, including emphysema, lung cancer, and fibrosis [4–7]. However, most of these studies rely on image analysis tools that require a great amount of user interaction or are merely qualitative.

Compared to manual or semi-automatic methods, fully automatic segmentation and quantification tools allow for accurate and reproducible measurements of disease progression. In this work, we apply these tools to a mouse model of chronic pulmonary inflammation caused by intratracheal instillation of silica. To the best of our knowledge, this is the first time that this model of inflammation is analyzed using micro-CT. Previously, Cavanaugh *et al.* had used micro-CT to quantify the extent of fibrotic and inflamed areas in a mouse model of lung injury [7]. The lung volume affected by fibrosis and inflammation was obtained by subtracting the healthy lung volume at baseline -areas of no X-ray absortion- from the healthy lung volume before the final image acquisition (a few days later). This approach assumes that the complete lung volume remains constant, which is not necessarily true. Lee *et al.* imaged the same animal model both in vivo and post mortem. The assessment was done by dividing the lung volume into 20 levels and visually assigning scores for different possible findings (honeycombing, ground-glass opacity, etc.). Generally, findings in post mortem scans correlated better with histology. Finally, Ask *et al.* studied a different fibrosis model in rats [8]. The fibrotic volumes were traced using region growing and compared to the histology-based fibrotic score.

Automatically segmenting inflammation areas in micro-CT images poses a considerable challenge. In some respects it is similar to the segmentation of computed tomography (CT) images of pathologies such as severe consolidation or extended fibrosis. Sluimer *et al.* worked on algorithms to segment scans with dense pathology [9] by combining image registration and voxel classification. More recently, Lee *et al.* used level sets to segment images of patients with diffuse interstitial lung disease (DILD) [10]. Their results were generally correct although the method failed in the most severe cases.

In this article, we propose a segmentation algorithm that combines threshold-based segmentation, atlas-based segmentation, and level sets. The method is validated by comparing computer-generated segmentations to manually-defined segmentations. The automatic segmentations of the lungs are used as masks to evaluate the extent of inflammation in a series of longitudinal scans. Quantitative measurements are performed on both diseased and healthy control mice. The rest of the article is organized as follows. In Section 2 the data acquisition method, the algorithms for lung segmentation and the disease quantification are explained. Section 3 presents the results of the segmentation evaluation, along with the disease quantification. Finally, in Section 4 some implications of the methods and results are discussed.

2 Methods

2.1 Data sets and image acquisition

We used a total of 40 A/J mice (Harlan UK Limited, Oxon, UK). The mice were 8 weeks old at the start of the experiments. Twenty-two mice were assigned to the control group and 18 to the silicosis group. Three different time points were defined for image acquisition: week 0 (before treatment), week 4, and week 14. In each group and time point, a number of animals was chosen for the follow-up study and the rest were sacrificed after the scan for purposes not relevant to this particular work. Table 1 shows the number of animals scanned in each time point.

Table 1: Animals scanned in each time point. (*) indicates that those animals were sacrificed after the scan. The rest made up the follow-up groups. Three mice of the silicosis group did not recover from the scanning protocol (one in time point 0 weeks, two in time point 4 weeks).

Time point	Control group	Silicosis group
Week 0	7 + 5*	8
Week 4	7 + 5*	7 + 5*
Week 14	7 + 5*	5 + 5*

Mice belonging to the control group were intratracheally instilled with 90 μl of saline. The silicosis group was instilled with crystalline silica (9 mg in 90 μl saline per mice). The crystalline silica sample was 99% pure alpha-quartz (Min-U-Sil 5; US Silica Co., Berkeley Springs, WV, USA), with a particle size of $< 5\ \mu m$. This single treatment causes silicosis which occurs with severe inflammation in the first weeks after instillation [11]. All procedures were carried out in compliance with European Union and University of Navarra (Institutional Animal Care and Use Committee) relevant guidelines for the use of laboratory animals.

Our imaging protocol is an adaptation of that by Namati *et al.* [12]. Animals were anesthetized with an intraperitoneal injection of 90 mg/kg ketamine and 10 mg/kg xylacine. Endotracheal intubation was performed on anesthesized animals using the BioLite system (Biotex, Houston, Texas), to illuminate the trachea with a fiber optic stylet. After intubation, animals were connected to a Flexivent rodent ventilator (Scireq, Montreal, Canada) at a rate of 200 breaths/min and a tidal volume of 10 ml/kg. Animals were kept breathing isoflurane at 2% concentration until complete relaxation was achieved and 0.5% isoflurane was maintained during the scan. 700 micro-CT projections were acquired during iso-pressure breath holds at 12 cm H_2O, which represents a physiological pressure and minimizes the probability of ventilator induced lung injury (VILI) [13]. Breath hold duration was 650 ms and normal breathing was induced during 1 s intervals between breath holds. Every 20 breath holds a total lung capac-

ity (TLC) perturbation was performed. This inflates the lung to the maximum capacity (30 cm H_2O) for 3 s and helps prevent atelectasis [14].

Images were acquired with a Micro-CAT II scanner (Siemens Pre-Clinical Solutions, Knoxville, Tennessee), with a source voltage of 80 kVp and a current of 500 μA. The exposure time per projection was 450 ms and each projection was acquired during the 650 ms iso-pressure apneae. This resulted in images of 640 slices with 1024×1024 voxels per slice, and an isotropic voxel size of 46 μm. An image of a water phantom was used as reference to convert the voxel values to Hounsfield Units (HU).

2.2 Lung segmentation

Due to the difficulty of the task, our lung segmentation algorithm combines three different segmentation strategies: threshold-based segmentation, atlas-based segmentation and geodesic active contours. This subsection details the implementation of each technique separately, concluding with the method used to combine the three segmentation results.

Threshold-based segmentation. We first segmented the healthy, i.e., low density, parts of the inflamed lungs using an algorithm largely based on previous work by Hu *et al.* [15]. Its basic steps are:

1. *Automatic thresholding*: a threshold is iteratively computed to separate air and body tissue in the scan.
2. *Background removal*: air regions connected to the borders are removed.
3. *Small volume removal*: small air volumes, which are likely to correspond to air blobs in the digestive tract, are removed.
4. *Hole filling*: possible holes within the lung are filled.
5. *Trachea extraction*: the trachea and the main airways are deleted from the volume.
6. *Small vessel inclusion*: a 3D closing operator is used to include small vessels in the lung segmentation.

With this method the largest connected area of the lung with low density is segmented, but the inflamed areas remain unsegmented.

Atlas-based segmentation. This step is essential for the correct inclusion of the diseased lung areas in the final segmentation. Furthermore, the result of this step will be used for the initialization of the final level set refinement. Atlas-based segmentation uses registration between a reference image, the atlas image, and the image to be segmented, the target image, to transform the segmentation of the atlas image into a segmentation of the target image [16].

In this case, a randomly chosen silicotic lung was manually segmented to serve as an atlas for the segmentation of the rest. Instead of directly registering the atlas and the target images, binary images containing rib segmentations were

used. The reason is that there is a large potential variability in inflammation location, extension and appearance, which can lead to errors in the registration. Although inflamed areas appear mainly near the main airways in the upper part of the lung, they can occasionally be in the lower parts too. Therefore, in all images ribs were segmented by thresholding and the resulting binary images were registered. A gaussian blurring was applied on the binarized images to help registration. The alignment process consisted of two steps: affine registration and B-Spline non-rigid registration. Mutual information was used as metric together with a stochastic gradient descent optimizer, which allows for fast and accurate registrations [17].

Geodesic Active Contours for refinement. The result of the atlas-based step is generally accurate at the ribs, has the expected lung shape in the upper, inflamed areas of the lung, but is inaccurate around the diaphragm and the heart. To refine the segmentation results in those areas we used geodesic active contours [18] . The geodesic active contours, which were implemented in ITK, follow the equation [19]:

$$\frac{d\psi}{dt} = -\alpha \mathbf{A}(\mathbf{x}) \cdot \nabla\psi - \beta P(\mathbf{x}) \cdot |\nabla\psi| + \gamma Z(\mathbf{x})\kappa \cdot |\nabla\psi| \quad (1)$$

where ψ is the level set function; α, β and γ are the weights for the advection, propagation and curvature terms, respectively; $\mathbf{A}(\mathbf{x})$ is an advection term; $P(\mathbf{x})$ is a propagation term; and $Z(\mathbf{x})$ is a spatial modifier for the curvature term κ.

To limit the refinement to the areas of interest near the diaphragm, we applied the active contour segmentation only to the lower two thirds of the lung in the coronal direction. The weights were set to 0.1 for propagation, 10 for curvature and 20 for advection. The high curvature and advection terms lead to smooth segmentations, which at the same time are strongly attracted to the edges.

Segmentation combination. The three previous segmentation results must be combined into a single final segmentation. In this fusion process, two basic assumptions are made:

- All voxels segmented by the threshold-based segmentation method or the level sets are correct.
- All voxels segmented by the atlas-based segmentation method are correct, as long as there is another lung voxel with the same x-y coordinates between the current point and the diaphragm, segmented by any of the other two methods. A x-y plane is defined as an axial or transverse slice of the micro-CT image.

With these assumptions, no extra voxel is added below the diaphragm, which is assumed to be well delineated by the combination of the threshold-based and the level sets segmentations. The atlas-based segmentation is used to add pathological areas int the upper area of the lung.

Figure 1 shows an example of the segmentation process.

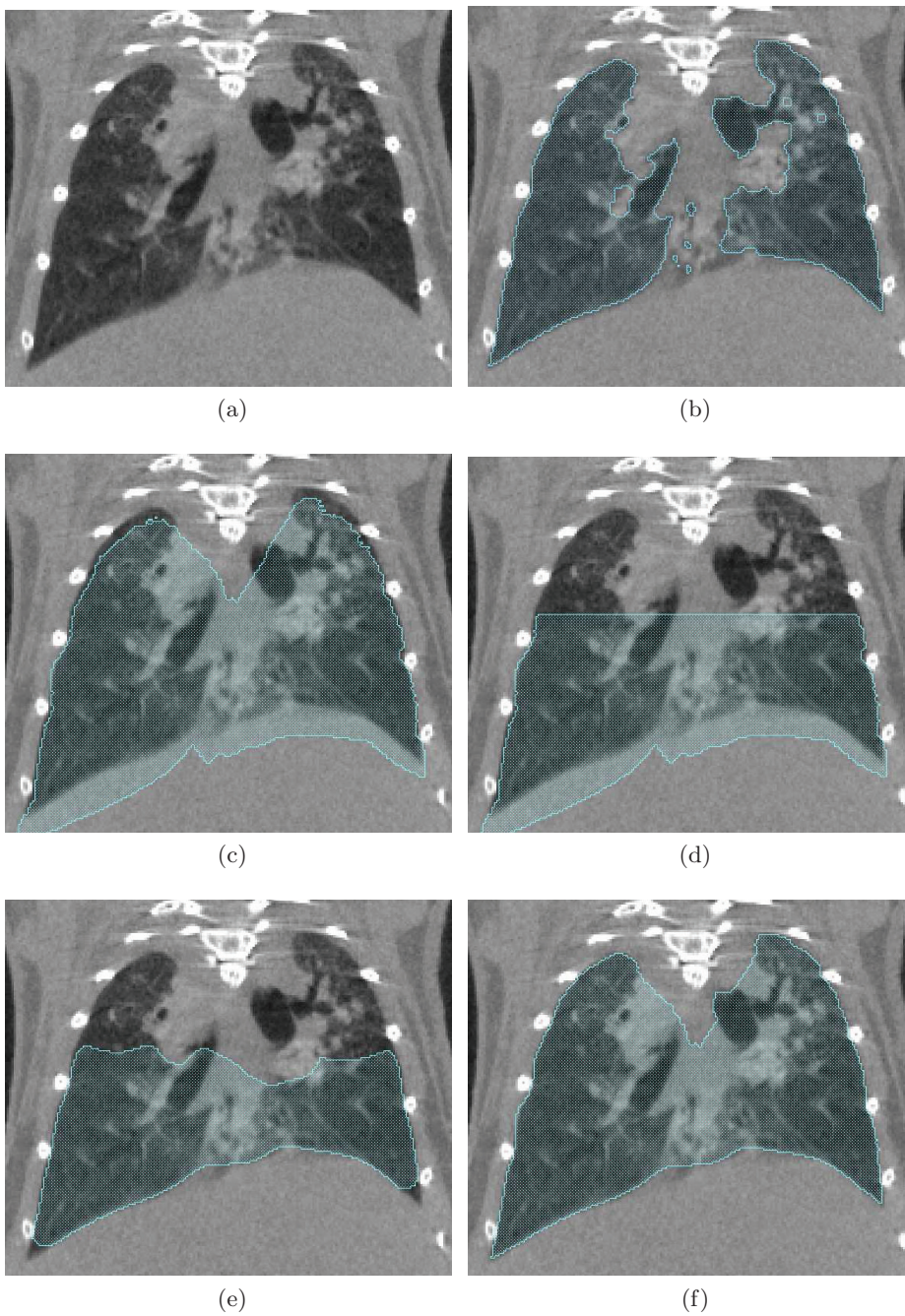


Fig. 1: (a) Micro-CT image of silicotic lung; (b) Threshold-based segmentation; (c) Atlas-based segmentation; (d) Cropped atlas-based segmentation; (e) Cropped atlas-based segmentation refined using level sets; (f) Final combined segmentation.

To segment healthy lungs, the atlas-based segmentation step was omitted. In principle, the threshold-based segmentation should suffice, but the level set method was used, with a dilated threshold-based segmentation as initial zero level set, as reported in [10]. This step was required to include the vessels in the segmentation, as occurs on the inflammation images.

In all cases an airway segmentation algorithm similar to the one reported by Artaechevarria *et al.* was used to remove the airways from the final lung segmentation [20].

3 Results

3.1 Segmentation evaluation

Six different scans were manually traced using the Amira software (Visage Imaging, Fürth, Germany) and were used as reference for the evaluation of the automatic segmentation method. The similarity index (SI) between the manual and the automatic segmentations was computed. The SI between two segmentations, S_a and S_b, of the same object, is defined as:

$$SI = \frac{2|S_a \bigcap S_b|}{|S_a| + |S_b|}, \tag{2}$$

where $\bigcap$ indicates the overlapping voxels between the two segmentations, and $|S_a|$ indicates the number of voxels of the corresponding segmentation [21]. SI has value 1 when there is a perfect match between labels and 0 when there is no overlap.

Table 2 shows the SI for the six manually segmented images. The average SI is 0.95.

Table 2: SI for the six manually segmented images

Image number and group	Similarity Index
Image 1 (control, week 0)	0.97
Image 2 (silica, week 0)	0.97
Image 3 (silica, week 4)	0.94
Image 4 (silica, week 4)	0.96
Image 5 (silica, week 14)	0.94
Image 6 (silica, week 14)	0.93

3.2 Disease quantification

Four different measurements were performed on the segmented lungs, in order to assess the effect of the inflammation:

1. *Total Lung Volume (TLV)*: corresponds to the final segmentation.
2. *Healthy Lung Volume (HLV)*: the result of the threshold-based segmentation.
3. *Inflamed Lung Volume (ILV)*: computed as TLV−HLV.
4. *Mean Lung Intensity (MLI)*: in HU, considering the complete lung.

Figure 2 shows the mean values of these parameters in control and diseased mice. There are no statistically significant differences between groups before treatment, but in weeks 4 and 14 silicotic mice show higher MLI, ILV and TLV. HLV is larger in the silica group only at week 4.

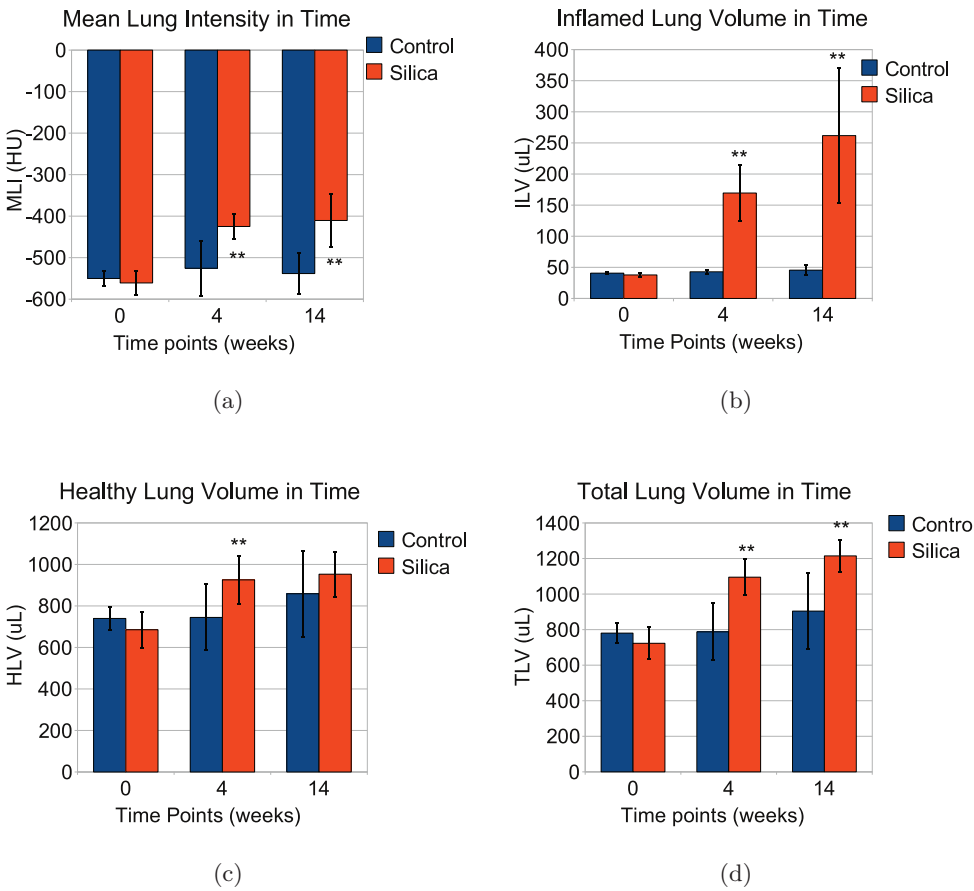


Fig. 2: (a) MLI vs. time points; (b) ILV vs. time points; (c) HLV vs. time points; (d) TLV vs. time points. (**) indicates $p < 0.05$, according to the MannWhitney U test. Error bars reflect the standard deviation within the group.

4 Discussion

The quantitative study of this model of lung disease using micro-CT requires accurate segmentations of the lungs. Automatic segmentation methods such as the one described in this work are essential, because manual segmentations are very time consuming, requiring as much as 5 hours in the diseased cases. Even though further evaluation is required, preliminary results show that our automatic segmentation method yields accurate results.

Atlas-based segmentation is one of the steps of the automatic segmentation method. A randomly chosen image of the silicosis group was chosen as the atlas image. Segmentation accuracy results show that this approach does not lead to large errors, but overall accuracy could improve with the use of a more generic atlas.

The calculated quantitative values are highly informative. As expected, we found a statistically significant increase of ILV and MLI in diseased versus to control animals. However, the higher HLV and TLV values found in silicotic mice compared to control ones requires further analysis. This result suggests that a compensatory mechanism exists, through which the healthy lung volume increases in response to the loss of functional tissue caused by inflammation. Further study, including histology and pulmonary functional tests, will be required to gain more knowledge on this phenomenon. It is also noteworthy that the mean ILV grows considerably within the diseased mice from week 4 to week 14, although a large variability exists.

In conclusion, we have shown that automatic image segmentation and quantification on a mouse model of lung inflammation can point at relevant biological issues. The intrinsic complexity of the segmentation task has produced the need for a custom segmentation method, within which several well-known methods are combined taking the characteristics of the expected disease into account.

References

1. de Torres, J.P., Bastarrika, G., Wisnivesky, J.P., Alcaide, A.B., Campo, A., Seijo, L.M., Pueyo, J.C., Villanueva, A., Lozano, M.D., Montes, U., Montuenga, L., Zulueta, J.J.: Assessing the Relationship Between Lung Cancer Risk and Emphysema Detected on Low-Dose CT of the Chest. Chest **132**(6) (2007) 1932–1938
2. Houghton, A.M., Mouded, M., Shapiro, S.D.: Common origins of lung cancer and COPD. Nature Medicine **14**(10) (2008) 1023–1024
3. Johnson, K.A.: Imaging Techniques for Small Animal Imaging Models of Pulmonary Disease: Micro-CT. Toxicologic Pathology **35**(1) (2007) 59–64
4. Recheis, W., McLennan, G., Ross, A.F., Hoffman, E.A.: Imaging the mouse lung with micro-CT. In: Molecular Imaging of the Lungs. Taylor and Francis (2005)
5. Froese, A.R., Ask, K., Labiris, R., Farncombe, T., Warburton, D., Inman, M.D., Gauldie, J., Kolb, M.: Three-dimensional computed tomography imaging in an animal model of emphysema. European Respiratory Journal **30**(6) (2007) 1082–1089

6. De Clerck, N., Meurrens, K., Weiler, H., Dyck, D.V., Greet, G.V., Terpstra, P., Postnov, A.: High-resolution X-ray microtomography for the detection of lung tumors in living mice. Neoplasia **6**(4) (2004) 374–379
7. Cavanaugh, D., Travis, E.L., Price, R.E., Gladish, G., White, R.A., Wang, M., Cody, D.D.: Quantification of bleomycin-induced murine lung damage in vivo with micro-computed tomography. Academic Radiology **13**(12) (2006) 1505 – 1512
8. Ask, K., Labiris, R., Farkas, L., Moeller, A., Froese, A., Farncombe, T., McClelland, G.B., Inman, M., Gauldie1, J., Kolb, M.R.: Comparison between conventional and "clinical" assessment of experimental lung fibrosis. Journal of Translational Medicine **6**(16) (2008)
9. Sluimer, I., Prokop, M., van Ginneken, B.: Toward automated segmentation of the pathological lung in CT. IEEE Transactions on Medical Imaging **24**(8) (2005) 1025– 1038
10. Lee, J., Seo, J.B., Kim, N., Park, S.O., Lee, H., Shin, Y.G., Kim, S.H.: Novel level-set based segmentation method of the lung at HRCT images of diffuse interstitial lung disease (DILD). In: SPIE Medical Imaging 2009: Image Processing. Volume 7259.
11. Saffiotti, U., Williams, A., Daniel, L., Kaighn, M., Mao, Y., Shi, X.: Carcinogenesis by crystalline silica: Animal, cellular, and molecular studies. In: Silica and silica-induced lung diseases. CRC Press (1996)
12. Namati, E., Chon, D., Thiesse, J., Hoffman, E.A., de Ryk, J., Ross, A., McLennan, G.: In vivo micro-CT lung imaging via a computer-controlled intermittent iso-pressure breath hold (IIBH) technique. Physics in Medicine and Biology **51**(23) (2006) 6061–6075
13. Dreyfuss, D., Saumon, G.: Ventilator-induced Lung Injury . Lessons from Experimental Studies. American Journal of Respiratory and Critical Care Medicine **157**(1) (1998) 294–323
14. Allen, G.B., Suratt, B.T., Rinaldi, L., Petty, J.M., Bates, J.H.T.: Choosing the frequency of deep inflation in mice: balancing recruitment against ventilator-induced lung injury. American Journal of Physiology - Lung Cellular and Molecular Physiology **291**(4) (2006) 710–717
15. Hu, S., Hoffman, E., Reinhardt, J.: Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. IEEE Transactions on Medical Imaging **20**(6) (2001) 490–498
16. Rohlfing, T., Brandt, R., Menzel, R., Russakoff, D.B., Maurer, Jr., C.R.: Quo vadis, atlas-based segmentation? In: The Handbook of Medical Image Analysis – Volume III: Registration Models. Kluwer Academic / Plenum Publishers (2005) 435–486
17. Klein, S., Staring, M., Pluim, J.: Evaluation of Optimization Methods for Non-rigid Medical Image Registration using Mutual Information and B-splines. IEEE Transactions on Image Processing **16**(12) (December 2007) 2879–2890
18. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. International Journal of Computer Vision **22** (1995) 61–79
19. Ibanez, L., Schroeder, W., Ng, L., Cates, J.: ITK Software Guide. Kitware (2005)
20. Artaechevarria, X., Muñoz-Barrutia, A., van Ginneken, B., Ortiz-de-Solórzano, C.: Fast murine airway segmentation and reconstruction in micro-CT images. In: SPIE Medical Imaging 2009: Biomedical Applications in Molecular, Structural, and Functional Imaging. Volume 7262.
21. Zijdenbos, A., Dawant, B., Margolin, R., Palmer, A.: Morphometric analysis of white matter lesions in MR images: Method and validation. IEEE Transactions on Medical Imaging **13**(4) (Dec 1994) 716–724

Assessment of Tracheal Stenosis Using Active Shape Models of Healthy Tracheas: A Surface Registration Study

Rômulo Pinho[1], Kurt Tournoy[2], Robert Gosselin[3], and Jan Sijbers[1]

[1] University of Antwerp, Department of Physics, VisionLab, Belgium
[2] Ghent University Hospital, Department of Respiratory Medicine, Belgium
[3] Ghent University Hospital, Department of Radiology, Belgium
{romulo.pinho, jan.sijbers}@ua.ac.be;
kurt.tournoy@ugent.be; robert.gosselin@uzgent.be

Abstract. Tracheal stenosis is a life threatening condition for which the successful treatment relies on the precise evaluation of its dimensions and severity. Recently, Active Shape Models (ASMs) were proposed for stenosis assessment and stent prediction, with promising results. An effective ASM, however, depends on the applied surface registration technique, which should not be influenced by the stenotic regions. The present work reviews previously proposed registration techniques and formulates a new method to estimate the shape of the healthy trachea of a patient with stenosis. Experiments with real and simulation data showed that the new method outperforms the conventional methods with respect to registration accuracy.

1 Introduction

Tracheal stenosis is a stricture of the windpipe that can be life threatening. Reconstruction or resection surgeries and the use of stents are important resources in the management of the condition. However, a successful treatment relies on the correct assessment of the stricture, which determines its location, length and degree of severity. Manual and computer aided approaches for the assessment of stenosis have been described in the literature [1–3].

Concomitantly, Active Shape Models (ASM) have been an important tool in computer aided diagnoses. In order to register the models to clinical data, the sum of the square of residuals between the model and the target is iteratively minimised. A problem may arise if the distribution of the residuals is not Gaussian, since standard least squares minimisation applied to non-Gaussian data distributions is known to be suboptimal. Deviations from Gaussian assumptions are normally evidenced by the presence of outliers in the data.

With the above concepts in mind, Pinho *et al.* claimed that correct assessment of stenosis depends on a good estimation of the healthy trachea of a patient and they used ASMs of tubular approximations of healthy tracheas for assessment of stenosis and prediction of stent dimensions [4]. The challenge in this

approach, however, is to avoid the influence of stenotic regions on the registration of the ASM to clinical data. For this, they used a method in which, at each iteration, residuals corresponding to shape landmarks over regions with stenosis have their influence on the registration reduced by keeping these landmarks fixed w.r.t. the shape obtained in the previous iteration. Eventually, the fixed landmarks act as a counterforce against the attraction of stenotic regions, aiding the model in producing the desired healthy trachea.

Despite the promising results, the work above lacked a thorough analysis of the surface registration mechanism and experiments with clinical data. In the present work, we build upon [4], concentrating on the surface registration step of the ASM, and add the following contributions:

- formulate *FixedLandmarks* as a new method to avoid the influence of misleading regions during the registration of ASMs;
- build the ASM using correspondence optimisation of landmarks, as proposed by Huysmans *et al.* [5], instead of the tubular approximations used in [4];
- investigate the behaviour of *FixedLandmarks* and other methods through a comprehensive set of experiments on clinical as well as on simulation data;
- present qualitative and quantitative comparisons between the registration methods and standard least squares with respect to the estimation of the healthy trachea.

We begin this paper with a review of ASMs and their use in the estimation of healthy tracheas, in Section 2. In Section 3, the registration mechanism and methods previously used with ASMs are briefly described before the formulation of the *FixedLandmarks*. In Section 4, the experiments and results are presented and the article is concluded in Section 5.

2 Active Shape Model of Healthy Tracheas

ASMs are built from a training set of N aligned shapes, $\mathbf{x}_i$, each represented by the concatenation of its n, d-dimensional landmarks, which must correspond across the training set. Principal Component Analysis then extracts the N eigenvectors and non-negative eigenvalues of the covariance matrix of the training set. New shapes $\mathbf{x}$ are obtained with a linear combination between the average shape of the training set, $\overline{\mathbf{x}}$, and the $dn \times N$ matrix of orthonormal eigenvectors, $\mathbf{P}$:

$$\mathbf{x} = \overline{\mathbf{x}} + \mathbf{P}\mathbf{b} \ , \tag{1}$$

where $\mathbf{b}$ is an $N \times 1$ vector of weights, which are the parameters of the model [6].

ASMs can be registered to an object of the class they represent by adjusting the parameter set $\mathbf{b}$. When the model is applied to an image, the registration is usually an iterative process: the landmarks of the shape generated by the model at the current iteration are moved along their normals, generating a candidate shape $\mathbf{y}$, which matches high gradients corresponding to edges of the target. Afterwards, a new set of parameters $\hat{\mathbf{b}}$ is computed in order to allow the model

to be registered to $\mathbf{y}$. The set $\hat{\mathbf{b}}$ which defines the best fit of the model to the candidate shape is obtained by minimisation of the squared error between $\mathbf{y}$ and $\mathbf{x}$, represented by the following error function:

$$\xi(\mathbf{b}) = (\mathbf{y} - \mathbf{x})^T(\mathbf{y} - \mathbf{x}) \ . \tag{2}$$

Expanding Eq. (2) with Eq. (1) and minimising ξ with respect to $\mathbf{b}$ results in:

$$\hat{\mathbf{b}} = \mathbf{P}^T(\mathbf{y} - \overline{\mathbf{x}}) \ . \tag{3}$$

This minimisation is herein referred to as *StandardLS*. This whole procedure is repeated until no significant changes have been made to the shape generated by the model at subsequent iterations.

Huysmans *et al.* proposed a method for shape modelling of cylindrical surfaces using cylindrical parametrisation [5]. In their method, the shapes of the training set are first aligned using the iterative closest point algorithm [7] and mapped on the unit cylinder, with a criterion to minimise distortions. The choice of landmarks along the boundaries of the shapes is made automatically, in the parametric domain. Likewise, the correspondences between the landmarks are established in the parametric domain, using minimum description length [8]. The optimised landmarks are later mapped back onto the original shapes. The boundary shapes, $\mathbf{x}_i$, are eventually described by the concatenation of n landmarks $\mathbf{x}_{\mathrm{v}_j} = (x_j, y_j, z_j)$ and used to build the model.

As shown in [6], shapes generated with ASMs resemble those in the training set. By constructing the model with healthy tracheas only, local distortions typical of stenotic geometry are not present. As a result, the edges in the image corresponding to regions with stenosis have low impact on local deformations of the ASM. Yet, the shape generated by the model can still be globally narrowed. In order to cope with this drawback, the registration is divided into two iterative stages. The first stage, a rigid registration, aligns the average shape of the model to the target trachea. This procedure aids the gradient based search in finding the location and orientation of the target trachea. In addition, landmarks of the shape generated by model which are located in the vicinity of regions with stenosis tend to remain far from their corresponding target. In the second, non-rigid registration stage, those landmarks are kept fixed at each iteration in order to minimise their influence on the adjustment of the model parameters. As the shape generated by the model iteratively deforms, the expected result is a trachea that matches the healthy regions of the target and produces an estimation for the healthy caliber of its narrowed parts.

3 Surface Registration

When the distribution of the residuals, $\{r_j | j = 1, \ldots, n\}$, between $\mathbf{x}$ and $\mathbf{y}$ in Eq. (2) is not Gaussian, due to the presence of outliers, *StandardLS* may produce suboptimal results. The literature presents different approaches to avoid the influence of outliers. In the remainder of this section, we review some of these approaches applied to ASMs and formulate a new one, called *FixedLandmarks*.

Weighted Least Squares The influence of outliers can be reduced by assigning weights to the contribution of each residual, modifying Eq. (2) to

$$\xi_w(\mathbf{b}) = (\mathbf{y} - \mathbf{x})^T \mathbf{W} (\mathbf{y} - \mathbf{x}) \ , \tag{4}$$

where $\mathbf{W}$ is a diagonal matrix of weights. Minimising ξ_w with respect to $\mathbf{b}$ yields

$$\hat{\mathbf{b}} = (\mathbf{P}^T \mathbf{W} \mathbf{P})^{-1} \mathbf{P}^T \mathbf{W} (\mathbf{y} - \overline{\mathbf{x}}) \ , \tag{5}$$

which is the basic formulation of weighted least squares (*WLS*) minimisation.

From the above definition, it is clear that a good choice of weights is key for an effective use of *WLS*. In the field of Robust Statistics, estimators that are less affected by deviations from Gaussian or other model assumptions can be devised to further improve the effects of *WLS* [9]. Rogers *et al.* used robust statistics with ASMs in different medical applications and filled matrix $\mathbf{W}$ with the Huber weighting function [10]. Theobald *et al.* compared several weighting functions for landmark occlusion detection, among which the Talwar, the Cauchy, and the Gaussian weighting functions performed best [11]. Fig. 1 shows the four functions described. In all of them, σ is the standard deviation of the residuals, which can be estimated at each iteration from the median of their absolute values [10].

$$w_{\text{huber}_i} = \begin{cases} 1, & r_i < \sigma \\ \sigma/|r_i|, & \sigma \le r_i < 3\sigma \\ 0, & r_i \ge 3\sigma \end{cases} \qquad w_{\text{talwar}_i} = \begin{cases} 1, & r_i < \sigma \\ 0, & r_i \ge \sigma \end{cases}$$

$$w_{\text{cauchy}_i} = \frac{1}{1 + \left(\frac{r_i}{\sigma}\right)^2} \qquad w_{\text{gauss}_i} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(r_i - \mu)^2}{2\sigma^2}}$$

Fig. 1. Huber, Talwar, Cauchy, and Gaussian weighting functions.

In the present work, we assume that the rigid registration stage of the ASM, as described in Section 2, results in a shape near healthy regions of the trachea and far from narrowed ones. Therefore, according to the definitions above, the r_i's corresponding to landmarks over these narrowed regions will be considered the outliers in the distribution.

Surface Extrapolation In this approach, the purpose is to use the model to predict missing parts of the target shape. At each iteration k of the registration, $\mathbf{y}^{(k)} \approx (\overline{\mathbf{x}} + \mathbf{P}\hat{\mathbf{b}}^{(k)})|_{\mathcal{L}}$, where $\mathcal{L}$, of size m, denotes the set of landmarks of the model actually used. It is possible that $m \ll n$, where n is the total number of

landmarks of the model. The parameter set $\hat{\mathbf{b}}^{(k)}$ is computed as in Eq. (2), but using only the components of $\mathbf{P}$ and $\mathbf{x}$ corresponding to the m target landmarks.

Rajamani *et al.* used extrapolation to predict the shape of the femur from manually sampled points during hip surgery [12]. In their method, the m sampled points are matched to the nearest landmarks of the shape generated by the model at each iteration. Furthermore, a weighting term added to the error function restricts the deformation freedom of the ASM as m decreases, forcing the model to produce shapes similar to the average shape. In [13], extrapolation was used to plan reconstructions of mandibular dysplasia. The ASM is registered to parts of the mandible that are considered as being regularly shaped.

Again, we assume that the shape generated by the rigid registration converges to a location near healthy regions of the trachea and far from those with stenosis. Thus, the set $\mathcal{L}$ will represent landmarks associated to healthy regions, to which the ASM is expected to yield the best possible match.

Fixed Landmarks Here we formulate a new registration technique, which we refer to as *FixedLandmarks*.

After the rigid registration, the landmarks of the shape generated by the model at the current iteration are displaced along their normals. If a high gradient is not found within a threshold distance $d > 0$, the corresponding landmarks remain fixed, while other landmarks are allowed to move as usual.

Let then $\mathbf{x}^{(k)} = \overline{\mathbf{x}} + \mathbf{P}\mathbf{b}^{(k)}$ be the shape generated with the model at any iteration k of the non-rigid registration. Let $\mathbf{y}^{(k+1)}$ be the candidate shape generated by displacing the landmarks of $\mathbf{x}^{(k)}$ and let $d\mathbf{y}^{(k+1)} = \mathbf{y}^{(k+1)} - \mathbf{x}^{(k)}$. As described in the previous paragraph, if $\|\mathbf{y}_{\mathrm{v}_j}^{(k+1)} - \mathbf{x}_{\mathrm{v}_j}^{(k)}\| > d$, then $\mathbf{y}_{\mathrm{v}_j}^{(k+1)} = \mathbf{x}_{\mathrm{v}_j}^{(k)}$ and $d\mathbf{y}_{\mathrm{v}_j}^{(k+1)} = \mathbf{0}$, where $j = 1 \ldots n$. In other words, some landmarks of the candidate shape $\mathbf{y}^{(k+1)}$ remain fixed w.r.t. $\mathbf{x}^{(k)}$. Grouping the $d\mathbf{y}_{\mathrm{v}_j}^{(k+1)} = \mathbf{0}$ and the corresponding columns of $\mathbf{P}^T$ results in two subsets of landmarks, $\mathcal{L}'$ and $\mathcal{L}''$, of sizes n' and n'', respectively, such that $d\mathbf{y}^{(k+1)}|_{\mathcal{L}''} = \mathbf{0}$. Let us then write

$$\hat{\mathbf{b}}^{(k+1)} = \mathbf{P}^T(\mathbf{x}^{(k)} + d\mathbf{y}^{(k+1)} - \overline{\mathbf{x}}) \ , \tag{6}$$

from Eq. (3), and split it into

$$\hat{\mathbf{b}}^{(k+1)} = \left[\mathbf{P}^T(\mathbf{x}^{(k)} + d\mathbf{y}^{(k+1)} - \overline{\mathbf{x}})\right]\Big|_{\mathcal{L}'} + \left[\mathbf{P}^T(\mathbf{x}^{(k)} + d\mathbf{y}^{(k+1)} - \overline{\mathbf{x}})\right]\Big|_{\mathcal{L}''} \ , \tag{7}$$

which does not affect the result. Since $d\mathbf{y}^{(k+1)}|_{\mathcal{L}''} = \mathbf{0}$, we finally obtain

$$\hat{\mathbf{b}}^{(k+1)} = \left[\mathbf{P}^T(\mathbf{y}^{(k+1)} - \overline{\mathbf{x}})\right]\Big|_{\mathcal{L}'} + \left[\mathbf{P}^T(\mathbf{x}^{(k)} - \overline{\mathbf{x}})\right]\Big|_{\mathcal{L}''} \ , \tag{8}$$

showing that $\hat{\mathbf{b}}^{(k+1)}$ is determined by both the displaced landmarks $\mathbf{y}^{(k+1)}|_{\mathcal{L}'}$ and the landmarks $\mathbf{x}^{(k)}|_{\mathcal{L}''}$, which remained fixed[1]. Consequently, when computing

$$\hat{\mathbf{x}}^{(k+1)} = \overline{\mathbf{x}} + \mathbf{P}\hat{\mathbf{b}}^{(k+1)} \ , \tag{9}$$

[1] Note that $\mathcal{L}'$ and $\mathcal{L}''$ can be different at each iteration.

$\hat{\mathbf{x}}^{(k+1)}$ will be the best fit, in a least squares minimisation sense, to $\mathbf{y}^{(k+1)}|_{\mathcal{L}'}$ and $\mathbf{x}^{(k)}|_{\mathcal{L}''}$. Provided that there are enough healthy areas around regions with stenosis, the fixed landmarks force the shape generated by the model to remain far from those regions, while enabling correct matches at the healthy areas. As the shape deforms iteratively, it progressively assumes the form of the desired healthy trachea, guided by the regions where correct matches occur.

4 Experiments

We carried out experiments on simulation as well as on clinical data in order to compare the different registration techniques discussed in Section 3. With the simulation data, ground truths were formally established and the experiments provided a reliable quantitative comparison between the registration techniques. The experiments with clinical data, in turn, provided a qualitative comparison between them.

To build the ASM, we used $N = 9$ healthy tracheas at total lung capacity, each with $n = 1024$ landmarks. The low-dose, chest CT scans of their respective patients were obtained from pulmonary medication studies carried out at the University Hospital of Antwerp, Belgium, and patient data were anonymised before the images were used. The tracheas were segmented from the images using a region growing algorithm dedicated to the segmentation of the airways [14] and then converted to a 3-dimensional shape with the marching cubes algorithm [15]. The 3-d shapes were later supplied to the correspondence optimisation algorithm, as in Section 2, which eventually produced the shapes used in the model.

4.1 Quantitative Comparison

In order to quantitatively compare the registration methods, we ran a large set of leave-one-out tests using simulation data. First, for each of the N healthy tracheas, 72 phantoms of stenosis – 24 anteriorly located (A), 24 posteriorly (P), and 24 roughly symmetrically narrowed (S) – were created. The stenotic areas were generated by applying a local erosion mask to the binary images of the segmented healthy tracheas until the stenosis achieved the desired shape [4]. The phantoms followed the categories of Fig. 2, based on [16], and were validated by an expert in the pulmonology field. For example, phantom I-1P represents a posterior stenosis of less than 25% along the upper third of the trachea.

For each run of the $T = N$ leave-one-out tests, the model was built with $N-1$ tracheas and was then registered to the phantoms created from the trachea not present in the training set. In all tests, the average shape of the ASM was initially roughly placed near the target trachea in the image. Moreover, the non-rigid registration methods were only triggered after the convergence of the rigid registration. Since the initial conditions were always the same, our evaluation is guaranteed to be fair.

The quality of the assessment of stenosis using the ASM strongly depends on the estimation of the healthy trachea of the patient. Therefore, the objective

of this set of experiments is to register the model to the phantoms using all the registration methods and to measure the distance between the estimated tracheas and their originally healthy counterparts. For this purpose, we employed the algorithm proposed in [17] to compute errors between surfaces using the Hausdorff distance.

The iteration limit for the registration was set to 200. The minimum squared error between shapes generated at subsequent iterations, i.e., $\xi_{\mathrm{r}}^{(k)} = (\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)})^T(\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)})$, was set to $10^{-7}\mathrm{mm}^2$. In the *FixedLandmarks*, the landmarks were displaced within a distance $d = 1$mm along their normals. For the *Surface Extrapolation*, the m landmarks of the candidate shape that guided the deformations were those that remained near the target surface ($d \leq 1$mm) after each iteration of the registration. In addition, we dropped the weighting term defined in [12], since shapes similar to the average shape of the model are very unlikely to produce a good estimation of a specific healthy trachea. In this way, the *Surface Extrapolation* becomes equivalent to using *WLS* with a step function yielding binary weights. Besides the robust approaches of Section 3, we also included in the comparison the *StandardLS*, i.e., not distinguishing between healthy and stenotic areas on the target surfaces. In total, we ran 4536 tests.

4.2 Qualitative Comparison

In addition to the simulation experiments, we made a retrospective study with chest CT scans from 3 patients. The use of the CT scans was approved by the ethics committee of the Ghent University Hospital (doc. ECUZG2009/140), Belgium, and patient data were anonymised before the images were used in the experiments. The 3 patients had stenosis with the following characteristics:

- Patient 1 had severe posterior stenosis along the lower half of trachea,
- Patient 2 had severe lateral stenosis along the two lower thirds of the trachea,
- Patient 3 had severe symmetrical stenosis along the lower half of the trachea.

No preprocessing was applied to any of the 3-d CT images. In addition, they were very anisotropic in the axial direction, with pixel resolution, in mm, (0.62, 0.62, 3.00), (0.98, 0.98, 5.00), and (0.44, 0.44, 3.00) respectively.

Since the registration is an iterative, edge based search, neighbouring organs and structures as well as noise may mislead the search. Therefore, the registration

CATEGORY	LOCATION AND LENGTH
I	Upper third of the trachea
II	Middle third of the trachea
III	Lower third of the trachea
I-II	Upper third extending to middle third
II-III	Middle third extending to lower third
I-III	Upper third extending to lower third

CATEGORY	DEGREE
1	<25%
2	26–50%
3	51–75%
4	>75%

Fig. 2. Categories of stenosis based on location and length (L) and degree (R).

is dependent on the initial search location. At this point, the initialisation of the registration is done manually, by conveniently placing the average shape of the model inside the image, namely, near the target trachea. For each patient, the initial position of the model was manually set for one registration method, recorded, and then replicated for all other methods. The results were reviewed by an expert in the pulmonology field in order to qualitatively compare all methods with respect to the estimated healthy trachea.

The ASM built for these experiments contained all the $N = 9$ healthy tracheas. As before, the maximum number of iterations was set to 200, the minimum $\xi_{\mathrm{r}}^{(k)}$ was set to $10^{-7}\mathrm{mm}^2$, and $d = 1\mathrm{mm}$. *WLS* and *StandardLS* were used in the same way as in the experiments with simulation data.

4.3 Results and Discussion

For the quantitative comparison between the registration methods using the simulation data, we subdivided the phantoms of each healthy trachea into $G = 10$ groups, according to the categories defined in Fig. 2, each with a different size S_g. The reasoning behind this subdivision is to show how the methods behaved relative to variations in location, length, and degree of stenosis across the whole set of T leave-one-out tests. Let us then define, for a test instance t, $\delta_{\max_{gtp}}$ and $\delta_{\mathrm{mean}_{gtp}}$ as the maximum and mean distances, respectively, between the estimated trachea for phantom p, of group g, and its original, healthy equivalent. As stated in Section 4.1, these distances are obtained using the algorithm proposed in [17]. Afterwards,

$$\overline{\delta}_{\max_{gt}} = \frac{1}{S_g}\sum_{p=1}^{S_g}\delta_{\max_{gtp}} \;\text{ and }\; \overline{\delta}_{\mathrm{mean}_{gt}} = \frac{1}{S_g}\sum_{p=1}^{S_g}\delta_{\mathrm{mean}_{gtp}} \tag{10}$$

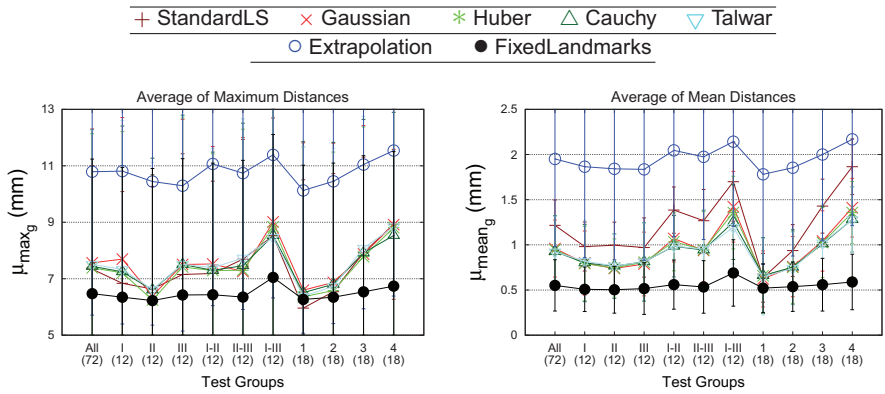
can be calculated as the average, per-group maximum and mean distances, respectively, for one test instance.

For the final comparison, we computed each group's average maximum error, $\mu_{\max_g}$, and average mean error, μ_{mean_g}, for each method, across the whole set of T leave-one-out tests. That is,

$$\mu_{\max_g} = \frac{1}{T}\sum_{t=1}^{T}\overline{\delta}_{\max_{gt}} \;\text{ and }\; \mu_{\mathrm{mean}_g} = \frac{1}{T}\sum_{t=1}^{T}\overline{\delta}_{\mathrm{mean}_{gt}} \;. \tag{11}$$

The results and their respective standard deviation bars are shown in Fig. 3.

As expected, the *StandardLS* did not perform well. The influence of the stenotic regions on the registration indeed made the resulting shape much narrower or deformed than desired. The use of *WLS* brought some improvements, but not enough to completely remove the influence of stenotic regions. The problem with the weighting approaches is the difficulty in finding the proper weight assigning function to act only on the regions with stenosis. If the weighting scheme is too tight, the shape may not deform enough, remaining similar to



(a) (b)

Fig. 3. Per-group $\mu_{\max_g}$ (a), μ_{mean_g} (b), and respective standard deviation bars for each method across the whole set of leave-one-out tests. Along the horizontal axis, the number of phantoms in the test group, S_g, is shown in parentheses.

the mean shape. If it is too loose, the shape may be strongly attracted by the areas with stenosis. Regarding the *Surface Extrapolation*, using only the points near the target surface to guide the deformations eventually resulted in few, very localised points, especially in the most severe cases. Without a stronger clue to indicate the shape to be obtained in a global level, the method could not converge to the desired result, which explains its poor performance. We can therefore conclude that the *FixedLandmarks* was the best registration method. It is especially worth noting how other methods performed worse as the length and degree of stenosis increased, while the *FixedLandmarks* was hardly affected. Its μ_{mean_g}'s remained near 0.5mm in all but one test group, I-III, which represented the longest and most severe types of stenosis in the simulation data.

Fig. 4 presents an example of the estimation of the healthy trachea for phantom II-4A generated from one of the healthy tracheas used in our experiments, using the *StandardLS*, *GaussWLS*, *HuberWLS*, *Surface Extrapolation*, and *FixedLandmarks* methods. The dashed, outermost silhouette in each case represents the original healthy trachea and the *FixedLandmarks* yielded the best fit to it.

As mentioned before, the results from the experiments with clinical data were reviewed by an expert in the pulmonology field. It can be seen in Fig. 5 that the *FixedLandmarks* produced very plausible healthy tracheas. They have an acceptable caliber and generally follow the curvature of the patient's trachea. Fig. 6 shows the results of other methods applied to the CT scan of patient 1. It can be seen how the severely narrowed trachea influenced the registration and either made the estimated tracheas too narrow and deformed or led them astray.

It is important to mention that the quality of the results obtained with the *FixedLandmarks* depends on the choice of the sets $\mathcal{L}'$ and $\mathcal{L}''$, which corresponds to landmarks that are allowed to move and those that remain fixed, respectively. These sets, in turn, depend on the choice of d. Intuitively, as d increases, the

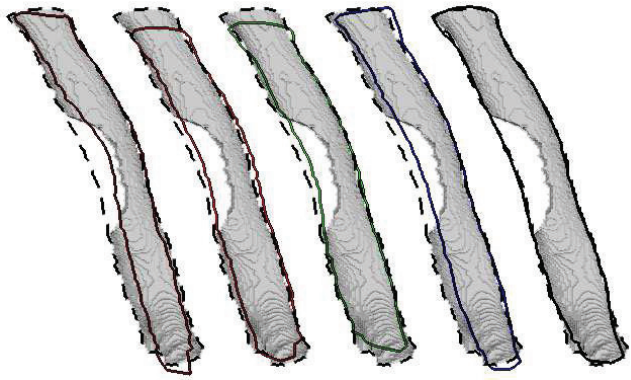


Fig. 4. Shape estimation for phantom II-4A of one healthy trachea from the simulation experiments, using, from left to right, the *StandardLS*, *GaussWLS*, *HuberWLS*, *Surface Extrapolation*, and *FixedLandmarks* methods. The dashed, outermost silhouettes represent the healthy trachea used to build the phantom.

FixedLandmarks tends to perform like the *StandardLS*, since $\mathcal{L}''$ will tend to be empty and no landmarks will remain fixed. The registration will thus not be guarded against the attraction of stenotic regions. If d is too short, $\mathcal{L}'$ will tend to be empty and, as opposed to the previous case, all landmarks will remain fixed. One option to solve this problem is to let this parameter be set by the user. Different values of d should then be tried until acceptable results are yielded. Another possibility is to devise an adaptive algorithm to change d as needed during the registration. Nevertheless, the value $d = 1$mm proved to be a good empirical choice in our comprehensive set of experiments.

The *FixedLandmarks* tended to fail when the rigid registration stage converged to a location where estimated healthy areas were either still far from the target trachea or too close to areas with stenosis. In the former case, the model shape was not attracted by the edges of the target. In the latter case,

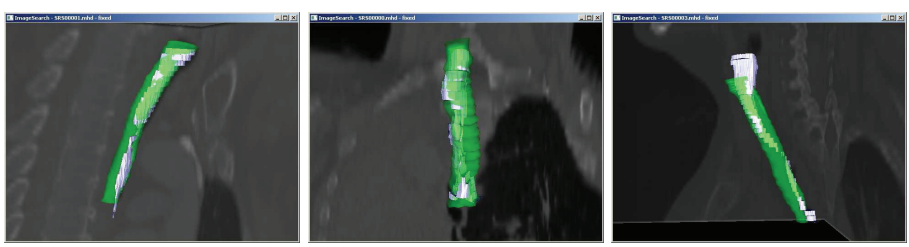


Fig. 5. Results of the estimation of the healthy trachea with the *FixedLandmarks* for patients 1, 2, and 3, from left to right. The estimated trachea, in green, is shown in the CT scan of the patient, overlaid on their segmented stenotic trachea.

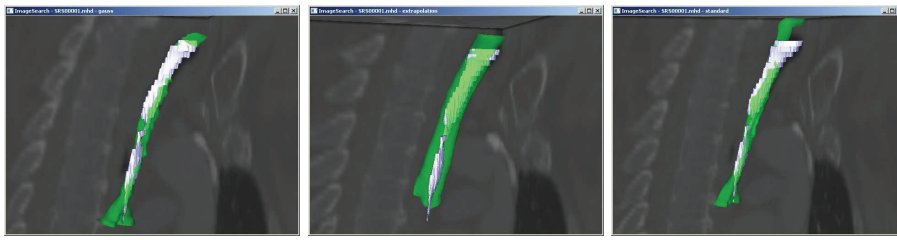


Fig. 6. *GaussWLS* (L), *Surface Extrapolation* (M), and *StandardLS* (R) registration methods applied to the CT scan of Patient 1. The estimated trachea, in green, is placed in the image, overlaid on the segmented stenotic trachea. With *Surface Extrapolation*, the estimated surface failed to match the lower part of the trachea. With other methods, the surface was too narrow or deformed.

the stenosis had stronger influence on the deformations, making the estimated shape somewhat narrower than desired. As stated above, the parameter d can be adjusted to try to reduce this problem, but further investigation of the rigid registration stage is still necessary.

Finally, we observed that problems with all approaches occurred mainly at areas where the shape of the trachea has more variation, namely the upper and lower thirds. This problem may be solved by an increase in the size and variability of the training set of the ASM and further experiments are ongoing[2].

5 Conclusion

We investigated the behaviour of registration methods used with Active Shape Models to estimate the healthy trachea of patients with tracheal stenosis. The estimated tracheas can be used, for instance, in surgery planning and prediction of stent dimensions. A new method, named *FixedLandmarks*, was formulated in order to avoid the influence of stenotic regions during the registration of the ASM to image data. The method works by keeping landmarks of the model associated to regions with stenosis fixed w.r.t. the previous iteration of the registration, which forces the shape generated by the model to stay far from these regions while enabling correct matches along healthy areas of the trachea. Experiments were carried out on simulation as well as on clinical data and the *FixedLandmarks* proved to be the best method when compared to other ones.

6 Acknowledgments

We thank the Pulmonology and the Radiology departments of the Ghent University Hospital and of the University Hospital of Antwerp for providing us with

[2] We refer the reader to http://www.youtube.com/user/fixedlandmarks for the latest results.

the images used in the experiments. We also thank the financial support of the IBBT (Interdisciplinary Institute for Broadband Technology), Belgium.

References

1. Carretta, A., Melloni, G., Ciriaco, P., Libretti, L., Casiraghi, M., Bandiera, A., ZanniniD, P.: Preoperative assessment in patients with postintubation tracheal stenosis. Surgical Endoscopy **20** (2006) 905–908
2. Kiesler, K., Gugatschka, M., Sorantin, E., Friedrich, G.: Laryngo-tracheal profile: a new method for assessing laryngo-tracheal stenoses. Eur Arch Otorhinolaryngol **264** (2007) 251–256
3. Shitrit, D., Valdsislav, P., Grubstein, A., Bendayan, D., Cohen, M., Kramer, M.: Accuracy of virtual bronchoscopy for grading tracheobronchial stenosis: correlation with pulmonary function test and fiberoptic bronchoscopy. Chest **128** (2005) 3545–3550
4. Pinho, R., Huysmans, T., Vos, W., Sijbers, J.: Tracheal stent prediction using statistical deformable models of tubular shapes. In: Proc. SPIE MI. (2008) 69144O
5. Huysmans, T., Sijbers, J., Vanpoucke, F., Verdonk, B.: Improved shape modeling of tubular objects using cylindrical parameterization. In: Proc. MIAR. Volume 4091. (2006) 84–91
6. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models: their training and application. Comput. Vis. Image Underst. **61** (1995) 38–59
7. Besl, P.J., McKay, N.D.: A method for registration of 3-d shapes. IEEE Trans. Pattern Anal. Mach. Intell. **14** (1992) 239–256
8. Davies, R.H., Twining, C.J., Cootes, T.F., Waterton, J.C., Taylor, C.J.: 3d statistical shape models using direct optimisation of description length. In: Proc. ECCV Part III. (2002) 3–20
9. Huber, P.J.: Robust Statistics. Wiley (1981)
10. Rogers, M., Graham, J.: Robust active shape model search. In: Proc. ECCV Part IV. (2002) 517–530
11. Theobald, B.J., Matthews, I., Baker, S.: Evaluating error functions for robust active appearance models. In: Proc. IEEE FG. (2006) 149–154
12. Rajamani, K.T., Styner, M.A., Talib, H., Zheng, G., Nolte, L.P., Ballester, M.A.G.: Statistical deformable bone models for robust 3d surface extrapolation from sparse data. Medical Image Analysis **11** (2007) 99–109
13. Zachow, S., Lamecker, H., Elsholtz, B., Stiller, M.: Reconstruction of mandibular dysplasia using a statistical 3d shape model. In: Proc. CARS. Volume 1281. (2005) 1238–1243
14. Pinho, R., Sijbers, J., Vos, W.: Efficient approaches to intrathoracic airway tree segmentations. In: Proc. IEEE EMBS Benelux Symposium. Volume 2. (2006) 151–154
15. Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3d surface construction algorithm. SIGGRAPH Comput. Graph. **21** (1987) 163–169
16. Freitag, L., Ernst, A., Unger, M., Kovitz, K., Marquette, C.: A proposed classification system of central airway stenosis. Eur Respir J. **30** (2007) 7–12
17. Aspert, N., Santa-Cruz, D., Ebrahimi, T.: Mesh: Measuring errors between surfaces using the hausdorff distance. In: Proc. IEEE ICME. Volume I. (2002) 705–708

Intensity-and-Landmark-Driven, Inverse Consistent, B-Spline Registration and Analysis for Lung Imagery

Kunlin Cao[1,3], Gary E. Christensen[1,3], Kai Ding[2,3], and Joseph M. Reinhardt[2,3,*]

[1] Department of Electrical and Computer Engineering
[2] Department of Biomedical Engineering
[3] The Iowa Institute for Biomedical Imaging and
The Iowa Comprehensive Lung Imaging Center
The University of Iowa, Iowa City, 52242

Abstract. Lung disease is the number three cause of death in America. Measuring local volume deformation from lung image registration may provide a non-invasive approach for detecting and classifying diseases and provide a means for measuring how these diseases respond to intervention. 3D lung CT images contain easily identifiable landmarks such as airway-tree and vascular-tree branch points that can be used for registration and validation. Intensity-based registration methods complement landmark registration methods by providing dense correspondence information since landmarks only provide sparse correspondence information. Intensity-based registration performs best in regions of strong contrast such as between the lung parenchyma and the chest wall, and between the parenchyma and the blood vessels and larger airways. This paper describes a Landmark, Inverse consistent, Tissue volume preserving, B-Spline (LITS) registration algorithm which can be used to measure local lung volume deformation. This method extends the original tissue volume preserving method by adding landmark information and inverse consistency constraint. LITS registration was applied on three subjects to match lung datasets acquired at functional residual capacity and total lung capacity. The registration errors are small compared to the large overall deformations. Sensitivity analysis was performed by changing node spacing of the parameterization and shows that finer B-Spline lattice setting can reveal more details of the feature deformation.

1 Introduction

The role of the respiratory system is to provide gas exchange. Lung tissue ventilation depends on the complex interrelationships between the lungs, rib cage, diaphragm, and abdomen. Many lung diseases, such as lung cancer and chronic obstructive pulmonary disease, alter the material properties of the lung tissue,

* Joseph M. Reinhardt is a shareholder in VIDA Diagnostics, Inc.

thus altering local and global region ventilation. Therefore, it is important to measure regional lung volume change for disease detection and tracking, and to help provide a better understanding of the normal and abnormal lung.

Imaging allows non-invasive study of lung behavior and image registration can be used to examine the lung deformation. Local lung expansion can be estimated using registration to match images acquired at different levels of inflation. Tissue expansion (and thus, specific volume change) can be estimated by calculating the Jacobian of the transformation [1]. It is important to ensure the accuracy of registration result since the Jacobian is computed from the transformation.

Guerrero et al. used two CT images acquired at different lung inflation levels and optical flow image registration to estimate regional ventilation to identify functioning vs. non-functioning lung tissue for radiotherapy treatment planning [2, 3]. Christensen et al. used consistent image registration to match images across cine-CT sequences to estimate local tissue expansion and contraction over the breathing cycle [4]. These methods assume that corresponding points in both images have the same grayscale intensities. However, this is not true for the lung because the tissue density changes as the lung inflates and deflates. CT intensity is a measure of tissue density and therefore changes as the tissue density changes. Efforts have been made to take this intensity change into consideration. Sarrut et al. [5] added a preprocessing step to artificially correct the intensity range. Gorbunova et al. [6] modified the sum of squared difference measure by adding total lung weight and regional volume change information to preserve lung weight globally and locally. Yin et al. [7] used a new similarity cost by preserving lung tissue volume. All these methods have demonstrated improved registration accuracy.

Most registration methods mentioned above use volumetric based similarity functions to compute a cost for mismatched subvolumes but give no cost to mismatched landmarks, contours, or surfaces. Mismatched landmarks, contours, and surfaces have zero measure under the volume integral and thus are difficult to match using a volumetric similarity function. Landmark based image registration has the advantage of matching landmarks either exactly or inexactly based on the confidence of the landmark location. A major limitation of landmark-based approaches is that they interpolate the transformation between the landmarks and ignore intensity information.

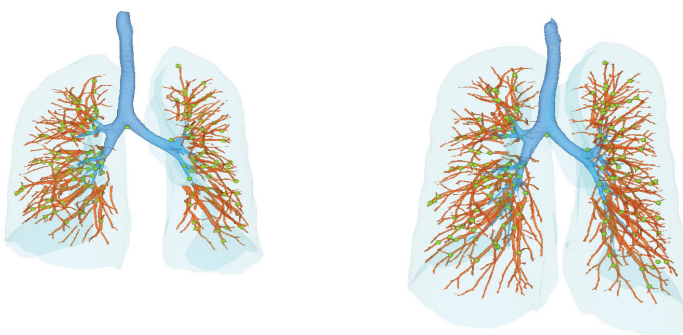
Combining landmark and intensity information together helps reduce the limitations of each method. This paper describes a Landmark, Inverse consistent, Tissue volume preserving, B-Spline (LITS) registration algorithm for matching lung CT images with large pressure changes. The sensitivity of this registration method to feature size is analyzed by changing the spatial resolution of the transformation parameterization.

2 Methods

2.1 Data Acquisition

Pairs of volumetric CT data sets from three normal human subjects scanned at supine orientation were used in the study. For subject 1, data was acquired at functional residual capacity (FRC) with 26.3% of the vital capacity (VC) and total lung capacity (TLC) with 95.7% of the VC. For subject 2, data was acquired at FRC with 21.8% of the VC and TLC with 95.6% of the VC. For subject 3, data was acquired at FRC with 11.0% of the VC and TLC with 68.9% of the VC. Volumetric data sets were acquired at a section spacing of 0.5 mm and a reconstruction matrix of 512 × 512. In-plane pixel spatial resolution is 0.6 mm × 0.6 mm. The data were resized to spatial resolution 1 mm × 1 mm × 1 mm to use less memory and reduce the time for registration.

The parenchyma regions in the FRC and TLC data sets were segmented using Hu et al [8]. An automatic lobe segmentation algorithm [9] was used to segment the parenchyma regions into five different lobes. Three fissures were identified where the lobe segmentations touched each other. The landmarks in FRC image were selected as the bifurcations of the vessel tree. A semi-automatic system [10] was used to guide the observer to find the corresponding landmarks in the TLC image. An expert selected 160, 150, and 105 landmark pairs for the three subjects, respectively. Fig. 1 shows the corresponding landmarks (green points) selected at vessel-tree branch points on FRC and TLC scans of one subject.



(a) (b)

Fig. 1. Landmarks selected at vessel-tree branch points on (a) FRC, and (b) TLC scans of one subject.

2.2 Inverse Consistent Image Registration

Let I_1 and I_2 denote two 3D images on the domain $\Omega = [0,1]^3$ and let $\mathbf{x} = (x_1, x_2, x_3)^T \in \Omega$ define a voxel coordinate in the domain Ω. The goal of a registration procedure is to find an optimal transformation (correspondence map) $\mathbf{h}$ that matches the corresponding coordinates of the two images. The transformation $\mathbf{h}$ is defined as a (3×1) vector-valued function defined on Ω. The vector-valued function $\mathbf{u} = \mathbf{h}(\mathbf{x}) - x$ is called the displacement field since it represents the transformation in terms of a displacement from a location $\mathbf{x}$.

The forward transformation $\mathbf{h_{12}}$ deforms I_1 to I_2 and the reverse transformation $\mathbf{h_{21}}$ deform I_2 to I_1. For intra-subject registration, a meaningful transformation should be one-to-one mapping, i.e., each point in image I_1 is mapped to only one point in image I_2 and vice versa. However, many unidirectional image registration techniques have the problem that their similarity cost function does not uniquely determine the correspondence between two images. The reason is that the local minima of similarity cost functions cause the estimated forward mapping $\mathbf{h_{12}}$ to be different from the inverse of the estimated reverse mapping $\mathbf{h_{21}^{-1}}$. To overcome correspondence ambiguities, the transformations $\mathbf{h_{12}}$ and $\mathbf{h_{21}}$ are jointly estimated in the LITS registration algorithm. Ideally, $\mathbf{h_{12}}$ and $\mathbf{h_{21}}$ should be inverses of one another, i.e., $\mathbf{h_{12}} = \mathbf{h_{21}^{-1}}$.

Depending on the application, different criteria for image matching are designed to find the best correspondence mapping between two images. When registering intra-subject CT images of the lung, many landmarks such as branch points of the airway and vascular trees can be easily identified and utilized for image registration and validation. The LITS registration algorithm is driven both by intensity and by minimizing the Euclidean distance between corresponding landmarks. The landmark error for a landmark on one image measures the distance from its estimated position to real position on the second image. A symmetric landmark similarity cost is defined to minimize the landmark errors,

$$C_{\mathrm{LMK}} = \sum_k ||\mathbf{p}_k - \mathbf{h_{12}}(\mathbf{q}_k)||^2 + ||\mathbf{q}_k - \mathbf{h_{21}}(\mathbf{p}_k)||^2, \tag{1}$$

where $\mathbf{p}_k$ and $\mathbf{q}_k$ are the location of landmark k on image I_1 and I_2, respectively.

The intensity matching criterion is used to register similar grayscale patterns in two images. To take the variation of intensity during respiration into account, the sum of squared tissue volume difference (SSTVD) [7] is used as the intensity similarity criterion to preserve tissue volume. This similarity criterion aims to minimize the local difference of tissue volume inside the lungs scanned at different pressure levels. Assume the Hounsfield units (HU) of CT lung images are primarily contributed by tissue and air. Then the tissue volume in a voxel at position $\mathbf{x}$ can be estimated as $V(\mathbf{x}) = v(\mathbf{x})\frac{HU(\mathbf{x})-HU_{air}}{HU_{tissue}-HU_{air}}$ where $v(\mathbf{x})$ is the volume of voxel $\mathbf{x}$. It is assumed that $HU_{air} = -1000$ and $HU_{tissue} = 55$. At location $\mathbf{x}$, let $I_1(\mathbf{x})$ and $I_2(\mathbf{x})$ be the intensity values (HU), $v_1(\mathbf{x})$ and $v_2(\mathbf{x})$ be the voxel volumes, and $V_1(\mathbf{x})$ and $V_2(\mathbf{x})$ be the tissue volumes in the voxel of

images I_1 and I_2, respectively. The intensity similarity cost is defined as

$$\begin{aligned} C_{\text{SSTVD}} &= \int_{\mathbf{\Omega}} \left\{ [V_2(\mathbf{x}) - V_1(\mathbf{h_{12}}(\mathbf{x}))]^2 + [V_1(\mathbf{x}) - V_2(\mathbf{h_{21}}(\mathbf{x}))]^2 \right\} d\mathbf{x} \\ &= \int_{\mathbf{\Omega}} \left\{ \left[v_2(\mathbf{x}) \frac{I_2(\mathbf{x}) + 1000}{1055} - v_1(\mathbf{h_{12}}(\mathbf{x})) \frac{I_1(\mathbf{h_{12}}(\mathbf{x})) + 1000}{1055} \right]^2 \right. \\ &+ \left. \left[v_1(\mathbf{x}) \frac{I_1(\mathbf{x}) + 1000}{1055} - v_2(\mathbf{h_{21}}(\mathbf{x})) \frac{I_2(\mathbf{h_{21}}(\mathbf{x})) + 1000}{1055} \right]^2 \right\} d\mathbf{x}. \quad (2) \end{aligned}$$

The Jacobian of a transformation estimates the pointwise volume change produced by mapping an image through the transformation. Thus, the tissue volume in image I_1 and I_2 are related by $v_1(\mathbf{h_{12}}(\mathbf{x})) = v_2(\mathbf{x}) \cdot J(\mathbf{h_{12}}(\mathbf{x}))$ and $v_2(\mathbf{h_{21}}(\mathbf{x})) = v_1(\mathbf{x}) \cdot J(\mathbf{h_{21}}(\mathbf{x}))$ where $J(\mathbf{h_{12}}(\mathbf{x}))$ and $J(\mathbf{h_{12}}(\mathbf{x}))$ are the Jacobian values of the forward and reverse transformations at position $\mathbf{x}$, respectively. In this way, the intensity similarity cost can be rewritten as

$$\begin{aligned} C_{\text{SSTVD}} = \int_{\mathbf{\Omega}} \left\{ \left[v_2(\mathbf{x}) \frac{I_2(\mathbf{x}) + 1000}{1055} - v_2(\mathbf{x}) J(\mathbf{h_{12}}(\mathbf{x})) \frac{I_1(\mathbf{h_{12}}(\mathbf{x})) + 1000}{1055} \right]^2 \right. \\ + \left. \left[v_1(\mathbf{x}) \frac{I_1(\mathbf{x}) + 1000}{1055} - v_1(\mathbf{x}) J(\mathbf{h_{21}}(\mathbf{x})) \frac{I_2(\mathbf{h_{21}}(\mathbf{x})) + 1000}{1055} \right]^2 \right\} d\mathbf{x}. \quad (3) \end{aligned}$$

Minimizing the symmetric cost functions given in Equations 1 and 3 does not guarantee that $\mathbf{h_{12}}$ and $\mathbf{h_{21}}$ are inverses of each other. In order to couple the estimation of $\mathbf{h_{12}}$ and $\mathbf{h_{21}}$, an inverse consistency constraint [11] is imposed which is minimized when $\mathbf{h_{12}} = \mathbf{h_{21}^{-1}}$:

$$C_{\text{ICC}} = \int_{\mathbf{\Omega}} ||\mathbf{h_{12}}(\mathbf{x}) - \mathbf{h_{21}^{-1}}(\mathbf{x})||^{\mathbf{2}} d\mathbf{x} + \int_{\mathbf{\Omega}} ||\mathbf{h_{21}}(\mathbf{x}) - \mathbf{h_{12}^{-1}}(\mathbf{x})||^{\mathbf{2}} d\mathbf{x}. \quad (4)$$

The transformation is said to be inverse-consistent when $\mathbf{h_{12}} = \mathbf{h_{21}^{-1}}$.

2.3 Transformation Parameterization and Estimation

The B-Splines based parameterization was used to represent the forward and reverse transformations to make it easy to combine landmark with an intensity based registration approach as shown by Kybic and Unser [12, 13]. Let $\phi_i = [\phi_x(\mathbf{x}_i), \phi_y(\mathbf{x}_i), \phi_z(\mathbf{x}_i)]^T$ be the i-th control point values on lattice G along each direction. The transformation is defined as

$$\mathbf{h}(\mathbf{x}) = \mathbf{x} + \sum_{i \in G} \phi_i \beta^{(3)}(\mathbf{x} - \mathbf{x_i}), \quad (5)$$

where $\mathbf{x}_i = (x_i, y_i, z_i)^T$ is the coordinate vector of a lattice point, $\beta^{(3)}(\mathbf{x}) = \beta^{(3)}(x)\beta^{(3)}(y)\beta^{(3)}(z)$ is a separable convolution kernel, and $\beta^{(3)}(x)$ is the uniform cubic B-Spline basis function.

The total cost $C_{total} = \alpha C_{LMK} + \rho C_{SSTVD} + \chi C_{ICC}$ is optimized using a limited-memory, quasi-Newton minimization method with bounds (L-BFGS-B) [14] algorithm, which is well suited for optimization with high dimensionality of parameter space. At each iteration, it was assumed that the functions $\mathbf{h_{12}^{-1}}(\mathbf{x})$ and $\mathbf{h_{21}^{-1}}(\mathbf{x})$ were constant to reduce computation complexity. α, ρ and χ are weights to adjust the significance of the three cost terms. For the study reported in this paper, the weights $\alpha = 4$, $\rho = 1$ and $\chi = 10^{-5}$ were selected by trial and error to give the best trade-off between intensity similarity and landmark similarity.

2.4 Assessment of Image Registration Accuracy

Point, surface, and volume measures were used to assess the accuracy of the image registration results. For each pair of data, 85% – 90% of landmarks were used for registration, and the remaining (15 landmarks) were used for validation. The landmark error measures the errors of validation landmarks, which are distributed in five lobes and were selected randomly.

The Fissure Positioning Error (FPE) is determined by comparing the distance between the transformed fissure and target fissure. The FPE is defined as the minimum distance between a point on the deformed fissure and the closest point on the corresponding target fissure. Mathematically, this can be stated as $FPE(\mathbf{x}) = \min_{\mathbf{y} \in F_2} d(\mathbf{x}, \mathbf{h_{12}}(\mathbf{y}))$ for a given point $\mathbf{x}$ in F_1, where F_1 (F_2, resp.) is the set of all points in the fissure in image I_1 (I_2, resp.) and $d(\cdot)$ defines the Euclidean distance.

The Relative Overlap (RO) statistic was used to measure how well corresponding regions of the parenchyma agreed with each other. The RO for forward registration is given by $RO(S_1 \circ \mathbf{h_{12}}, S_2) = \frac{(S_1 \circ \mathbf{h_{12}}) \cap S_2}{(S_1 \circ \mathbf{h_{12}}) \cup S_2}$ where S_1 and S_2 are segmentations of parenchyma regions on images I_1 and I_2, respectively. $RO = 1$ implies a perfect segmentation matching.

3 Experiments and Results

3.1 Registration Accuracy

The LITS registration method described in Section 2 was used to register the parenchyma region of the three subjects. Small Deformation Inverse Consistent Linear Elastic [11] (SICLE) registration, which uses the sum of squared difference as similarity cost and its transformation represented by Fourier coefficients, was used to register the three pairs of data for comparison. The mean and standard deviation of different errors for both forward (FRC to TLC) and reverse (TLC to FRC) transformations are shown in Fig. 2 (a)-(c). For this study, the SICLE algorithm used $3 \times 24 \times 24 \times 24 = 41,472$ parameters and LITS estimated $3 \times 25 \times 25 \times 25 = 46,875$ parameters (16 mm grid spacing) to parameterize each transformation. The average distance of validation landmarks across three subjects was 25.18 mm before registration (15 validation landmarks for each

subject). The SICLE algorithm reduced the landmark error to 7.53 mm and the LITS algorithm reduced it to 1.26 mm. The average fissure positioning error across three subjects was 8.22 mm before registration, 4.05 mm after using SICLE registration and 1.65 mm after using LITS registration. Both SICLE and LITS achieve around 0.96 relative overlap.

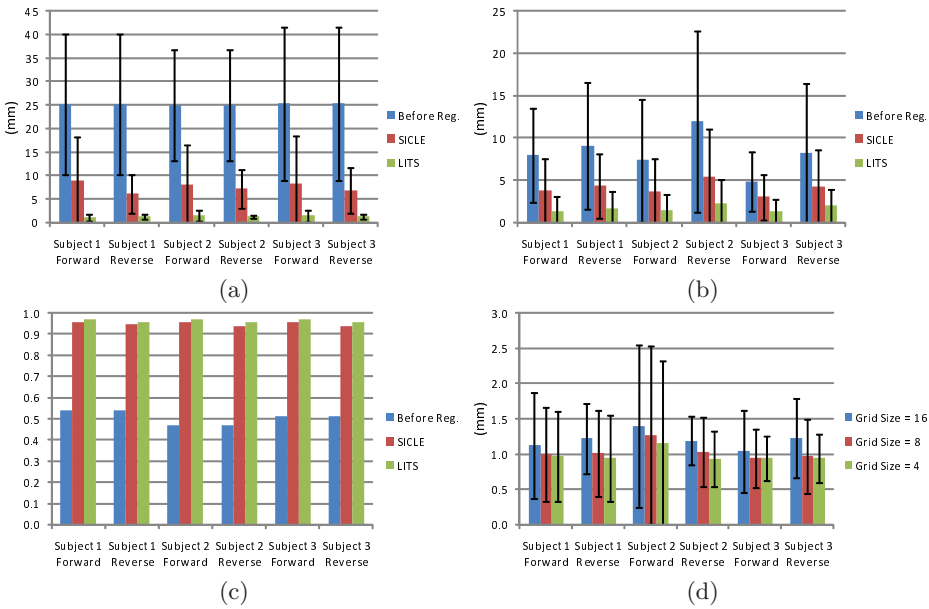


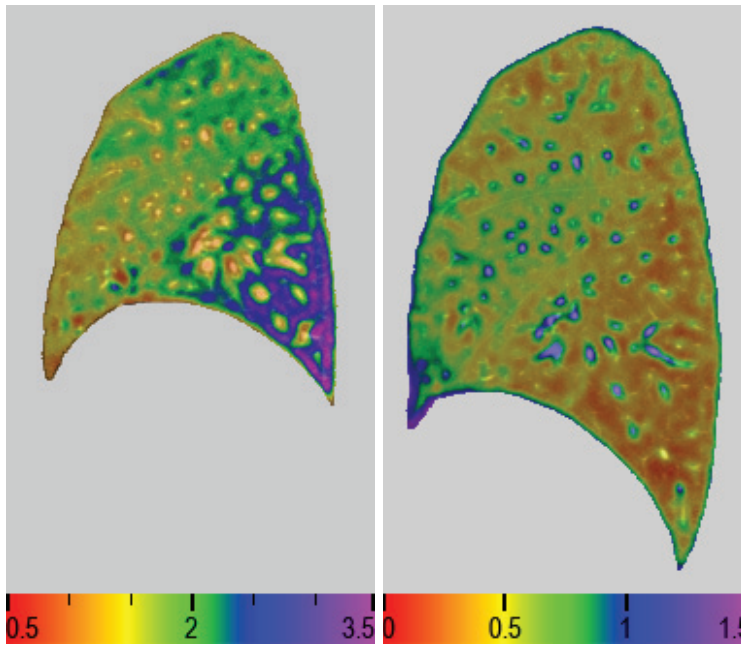
Fig. 2. The registration errors of experiments on three subjects. (a)-(c) show the mean and standard deviation for landmark error, fissure positioning error, and relative overlap using SICLE and LITS, respectively. (d) shows the mean and standard deviation of landmark error using LITS by varying the grid sizes.

3.2 B-Spline Sensitivity Analysis

The grid size of the B-Spline lattice decides the number of parameters estimated in the registration process. Three choices of grid size (4 mm, 8 mm and 16 mm) were tested using LITS to study how the grid size affected the registration results. The fissure positioning error and relative overlap error were nearly the same for different grid size. Fig. 2 (d) shows the landmark accuracy using different grid sizes. Smaller grid sizes have slightly lower errors.

The Jacobian of a transformation can be used to estimate the pointwise volume expansion and contraction of the transformation. Using a Lagrangian reference frame, local tissue expansion corresponds to a Jacobian greater than one and local tissue contraction corresponds to a Jacobian less than one. The Jacobian map shown in Fig. 3 reflects the fact that vessels have little deformation

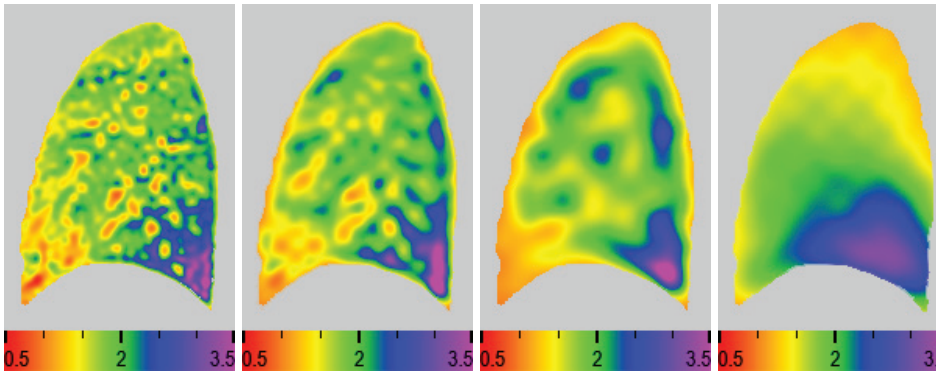
during the respiratory cycle while lung tissues and airways expand/contract a lot.



(a) (b)

Fig. 3. The Jacobian maps on a sagittal slice. (a) Jacobian on FRC image showing local volume deformation in inhalation stage; and (b) Jacobian on TLC image showing local volume deformation in exhalation stage. The B-Spline lattice grid size is 4 mm for LITS.

The Jacobian difference between vessels and other lung regions can be utilized to reveal how small a structure can be detected as the grid size of the B-Spline lattice is changed. Using LITS with different lattice grid size settings and SICLE, the Jacobian maps on the same sagittal slice of a FRC data are compared in Fig. 4. Fig. 5 shows the effect of changing grid size to detect vessels of different sizes and to detect two vessels in real images based on LITS resulting Jacobians (FRC to TLC). The Jacobian of the transformation produced using SICLE is also plotted for comparison. Each picture in Fig. 5 plots the 1/Jacobian (left vertical axis) and HU value (right vertical axis) changes along a line profile. The horizontal axis lists the point numbers on the line profile.



(a) (b) (c) (d)

Fig. 4. The Jacobian maps on the same sagittal slice of a FRC data resulting from LITS with the B-Spline lattice grid size of (a) 4 mm, (b) 8 mm, and (c) 16 mm. (d) shows the Jacobian map resulting from SICLE.

4 Discussion

Fig. 2 (a)-(c) shows that both SICLE and LITS match the boundaries of two lung parenchyma well but LITS has much lower landmark error and fissure positioning error compared to SICLE. There are two reasons for this. First, LITS has vessel tree landmark information in the registration while SICLE does not. Secondly, SICLE uses the sum of squared difference as similarity cost which assumes the grayscales for corresponding voxels in two images are the same. However, the assumption is not true in the lung CT images due to the tissue density change as air goes into the lung during inhalation. Taking this into consideration, the SSTVD similarity cost aims to minimize the lung tissue volume change during the respiratory process. This similarity cost performs better on vessels and fissures for intra-subject lung CT image registration.

Changing the grid size of the B-Spline lattice in LITS has little effect on the landmark error, fissure positioning error, and relative overlap measurements. This is because those measures evaluate the accuracy on mostly vessel branch points, fissures and boundaries which have strong intensity contrast with their neighboring region and can be matched through optimizing the intensity and landmark similarity cost. But finer grid setting can give a more detailed deformation map, as shown in Fig. 4 (a)-(c). Fig. 4 (d) shows that the 3D Fourier basis parameterization that used the 24 lowest harmonics in the 3 coordinate directions ($3 \times 24 \times 24 \times 24$ parameters) produced a smoother Jacobian map than the B-Spline basis. This result may suggest that the local support of the B-Spline basis is better for detecting small features compared to the infinite support of the Fourier series basis. Fig. 5 shows that smaller grid sizes are required to detect narrow vessels and to separate nearby vessels. This analysis can help us chose the B-Spline lattice spacing setting to detect the deformation of features with

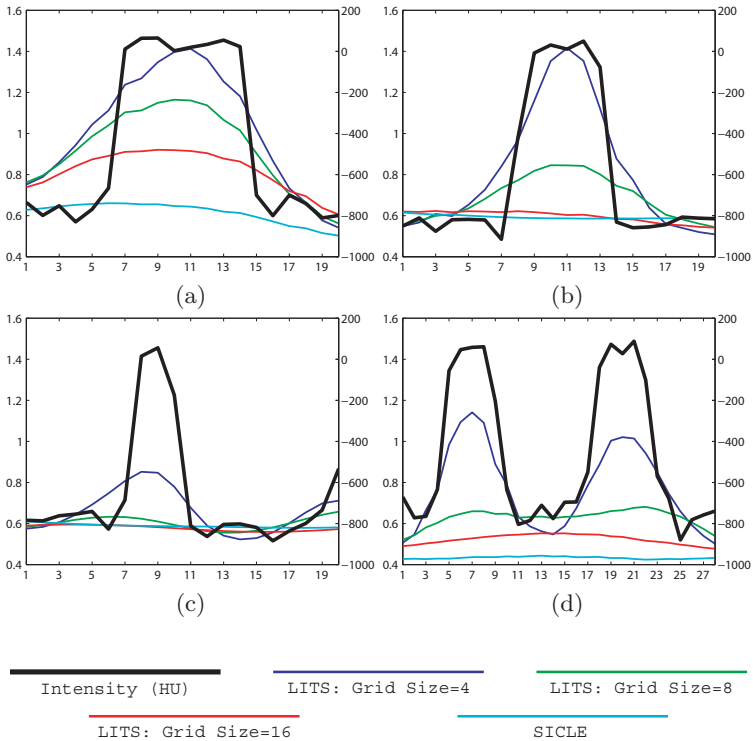


Fig. 5. The changes of 1/Jacobian (left vertical axis) and HU value (right vertical axis) on line profiles (a)-(c) across vessels of different sizes, and (d) across two nearby vessels. The horizontal axis lists the point numbers on the line profiles. Jacobian values are estimated from SICLE resulting transformation, and from LITS resulting transformations using the B-Spline lattice grid size of 4 mm, 8 mm and 16 mm.

acceptable accuracy and time cost. In comparison, the Jacobian from SICLE is too smooth to be used to detect the deformation of small local features.

The Jacobian of the transformation can be used to measure local volume deformation. Fig. 3 shows that vessels have much smaller volume changes comparing with parenchyma tissue and airways during breathing cycle. In addition, since the CT images were acquired with subjects in the supine orientation, an obvious dorsal to ventral gradient is noticed on Fig. 3 and Fig. 4. This is consistent with the well known physiology that subjects positioned in the supine posture have more ventilation in the dorsal region than in the ventral region.

5 Conclusion

We described the LITS image registration method that utilizes both intensity and landmark information to register 3D CT images of lung at different pressures. Forward and reverse transformations between two images were jointly

estimated to minimize the inverse consistency error. Landmarks were selected at major vascular tree bifurcations. They were used partly for registration and partly for validation. A recently proposed intensity similarity metric SSTVD [7] was used to ensure the parenchyma tissue volume preserves during breathing cycle. The registration accuracy of LITS evaluated through measurements of landmark error, fissure positioning error, and relative overlap were better than that of SICLE which proves that this landmark-and-intensity driven method can match two lung images with large deformations effectively.

In the future, we plan to combine landmark, contour, surface and intensity based image registrations to minimize registration artifacts that negatively impact analysis. Contours such as the centerline of airways and vessels can be registered by matching landmarks defined at equally spaced intervals along the contour. Our hypothesis is that a method that uses all the correspondence information regarding the lung image registration task will produce a better result than any single type of information [15].

6 Acknowledgments

The authors would like to thank Youbing Yin for his assistance of understanding and implementing the tissue volume preserving measure. This work was supported in part by the NIH grants HL079406, HL64368, HL080285, CA129022, and EB004126.

References

1. Reinhardt, J.M., Ding, K., Cao, K., Christensen, G.E., Hoffman, E.A., Bodas, S.V.: Registration-based estimates of local lung tissue expansion compared to xenon-CT measures of specific ventilation. Medical Image Analysis **12**(6) (2008) 752–763 DOI: 10.1016 / j.media.2008.03.007.
2. Guerrero, T., Sanders, K., Noyola-Martinez, J., Castillo, E., Zhang, Y., Tapia, R., Guerra, R., Borghero, Y., Komaki, R.: Quantification of regional ventilation from treatment planning CT. Int. J. Radiation Oncology Biol. Phys. **62**(3) (2005) 630–634
3. Guerrero, T., Sanders, K., Castillo, E., Zhang, Y., Bidaut, L., Komaki, T.P.R.: Dynamic ventilation imaging from four-dimensional computed tomography. Phys Med Biol. **51**(4) (2006) 777–791
4. Christensen, G.E., Song, J.H., Lu, W., Naqa, I.E., Low, D.A.: Tracking lung tissue motion and expansion/compression with inverse consistent image registration and spirometry. Med Physics **34**(6) (2007) 2155–2165
5. Sarrut, D., Boldea, V., Miguet, S., Ginestet, C.: Simulation of four-dimensional ct images from deformable registration between inhale and exhale breath-hold CT scans. Medical Physics **33**(3) (2006) 605–617
6. Gorbunova, V., Lo, P., Ashraf, H., Dirksen, A., Nielsen, M., de Bruijne, M.: Weight preserving image registration for monitoring disease progression in lung CT. In: MICCAI 2008. Volume LNCS 5242., Springer (2008) 863–870
7. Yin, Y., Hoffman, E.A., Lin, C.L.: Mass preserving non-rigid registration of CT lung images using cubic B-spline. Medical Physics **36**(9) (2009) In press.

8. Hu, S., Hoffman, E.A., Reinhardt, J.M.: Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. IEEE Trans. on Medical Imaging **20** (2001) 490–498
9. Ukil, S., Reinhardt, J.M.: Anatomy-guided lung lobar surface detection in X-ray CT images. IEEE Trans. Medical Imaging **28**(2) (2009) 202–214
10. Murphy, K., van Ginneken, B., Pluim, J., Klein, S., Staring, M.: Semi-automatic reference standard construction for quantitative evaluatoin of lung CT registration. In: Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention 2008. Volume 5242. (2008) 1006–1013
11. Christensen, G., Johnson, H.: Consistent image registration. IEEE Trans. Med. Imaging **20**(7) (2001) 568–582
12. Kybic, J.: Elastic Image Registration Using Parametric Deformation Models. PhD thesis, Swiss Federal Institute of Technology Lausanne, EPFL, Lausanne, Switzerland (2001)
13. Kybic, J., Unser, M.: Fast parametric elastic image registration. IEEE Transactions on Image Processing **12**(11) (2003) 1427–1442
14. Byrd, R.H., Lu, P., Nocedal, J., Zhu, C.: A limited memory algorithm for bound constrained optimization. SIAM J. Sci. Comput. **16**(5) (1995) 1190–1208
15. Miller, M., Banerjee, A., Christensen, G., Joshi, S., Khaneja, N., Grenander, U., Matejic, L.: Statistical methods in computational anatomy. Statistical Methods in Medical Research **6** (1997) 267–299

Quantification of Nodule Detection in Chest CT: A Clinical Investigation Based on the ELCAP Study

Amal A. Farag, Shireen Y. Elhabian, Salwa A. Elshazly and Aly A. Farag
Computer Vision and Image Processing Laboratory
University of Louisville, Kentucky, USA

Abstract. This paper examines the detection step in automatic detection and classification of lung nodules from low-dose CT (LDCT) scans. Two issues are studied in detail: nodule modeling and simulation, and the effect of these models on the detection process. From an ensemble of nodules, specified by radiologists, we devise an approach to estimate the gray level intensity distribution (Hounsfield Units) and a figure of merit of the size of appropriate templates. Hence, a data-driven approach is used to design the templates. The paper presents an extensive study of the sensitivity and specificity of the nodule detection step, in which the quality of the nodule model is the driving factor. Finally, validation of the detection approach on labeled clinical dataset from the Early Lung Cancer Action Project (ELCAP) screening study is conducted. Overall, this paper shows a relationship between the spatial support of the nodule templates and the resolution of the LDCT, which can be used to automatically select the template size. The paper also shows that isotropic templates do not provide adequate detection rate (in terms of sensitivity and specificity) of vascularized nodules. The nodule models in this paper can be used in various machine learning approaches for automatic nodule detection and classification.

Keywords: Parametric Templates Matching, Lung Nodule Definitions, Sensitivity and Specificity of CAD systems.

1 Introduction

The goal of computer-based nodule analysis methods is to assist the radiologists in early detection of presumable nodules. Assistance means to be able to mimic what a physician does in detecting and judging doubtful nodules. In the United States, lung cancer accounts for over 30% of all cancer–related deaths, resulting in over 160,000 deaths per year [1]. That is more than the annual deaths from colon, breast, pancreatic, prostate, and ovarian cancers combined. Lung cancer survival is strongly dependent on the pathologic stage at the time of diagnosis [2][3]. The hope is that early detection of lung cancer can improve the survival rate of this disease, thus research studies to reach an optimal detection rate is important. Should the use of LDCT scans become a standard clinical practice (for example, as a component of annual physical exams), an automatic way to analyze the scans will lend great benefit for the entire healthcare system; e.g., [4]-[7] and extensive surveys in [8][9].

The generalized framework for a CAD system consists of four main steps (see, Figure 1): 1) Filtering and normalization of the LDCT scans. 2) Segmentation of the lung regions (parenchyma) from the surrounding tissue. 3) Detection of lung nodules and 4) Nodule classification as benign or malignant.

Fig. 1: A block diagram of the major steps involved in computer-based analysis of LDCT of the chest in order to detect and classify doubtful lung nodules.

This paper will focus on the third step; nodule detection. The ELCAP dataset [10] is used for nodule design and testing and the sensitivity and specificity of the template matching approach in terms of detection is studied.

Since the early 90's various approaches for automated pulmonary nodule detection have been introduced. These approaches can be categorized as follows [4]-[7]: model based and density-based approaches. Template matching is one technique for model-based approaches which utilize a priori information of the size, intensity and shape of the nodules. Density-based approaches uses the fact that the lung parenchyma has relatively lower CT attenuation (density) than those of the lung nodules, thus they utilize image processing algorithms that rely on multiple thresholding, region growing and clustering. Various other approaches from the computer vision literature have been employed as well (e.g. [14]), which will not be surveyed due to space limitations.

Since the components of the CAD system in Fig. 1 are serial, we will briefly mention the techniques we employed in the filtering of scan artifacts, and segmentation of the lung tissues. For extensive surveys of various approaches on LDCT CAD research, the reader is directed to [8][9].

This paper is organized as follows: section 2 briefly discusses filtering of the scanning artifacts and segmentation of the lung region from the surrounding tissues; section 3 discusses template modelling and simulation; section 4 discusses template matching; section 5 presents experimental results and validation; and section 6 concludes the paper.

2 Artifacts Removal and Lung Segmentation in LDCT Scans

Filtering: During the acquisition of CT scans the generated slices can be affected by contrast, resolution, noise, artifacts and/or distortion. In particular, a slice can be corrupted by random fluctuations in image intensity which is considered a source of noise, thus, a CT slice is required to be filtered prior to further processing such as

segmentation. In this work we use the anisotropic diffusion filter (e.g., [11]), which is a well proven approach in various image analysis applications.

Image segmentation: as the name implies, connotes dividing or separating the information content in an image (or volume of images) into recognizable classes. It is a very fundamental and important step in image analysis. Various approaches have been used to segment the lungs from the rest of the organs that show the chest CT scans; again, the reader is referred to the surveys in [8][9]. The segmentation algorithm employed in this paper (similar to Armato et al [5]) exploits the intensity characteristics of lung CT scans, which appears as two distinct modes in the gray level intensity (Hounsfield Units) histogram. Multi-level thresholding is used to first isolate the thoracic region from the CT slice background; the lung parenchyma is then extracted from the segmented thoracic region, Fig. 2.

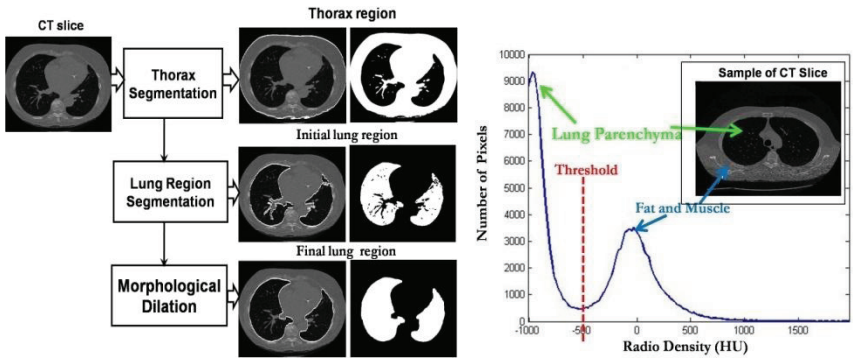


Fig. 2: *A Block diagram of the segmentation algorithm. Lung region (parenchyma) and fat/muscle region constitute two dominant peaks in the histogram, in order to separate the lung region a threshold is chosen to maximize the separation.*

Morphological dilation is then applied as a smoothing filter on the contour of the segmented lung region in order to avoid losing nodules that may be attached to the lung walls. To decrease the sensitivity of the segmentation result to the structuring element diameter, we apply it to the inner and outer lung region contour. After segmentation was completed small nodules attached to the pleural surface were found to no longer exist since these nodules were segmented as not belonging to the lung parenchyma. This operation resulted in 6.5% of the ground truth nodules to be excluded from further experimentations. Various improvements to the segmentation are possible; the algorithm employed here is 93.5% accurate with respect to retaining the nodules.

3 Nodule Modeling and Simulation

Proper nodule modeling is crucial for successful detection and classification. Modeling involves the shape, spatial support and the appearance (intensity) of the

template. The focus of this paper is on data-driven approaches for nodule modeling and simulation.

3.1 Nodule Types

A pulmonary nodule usually has a spherical shape; however, it can be distorted by surrounding anatomical structures such as vessels and the pleural surface. We shall use the nodule classification of Kostis et al [12][13], which groups nodules into four categories:

i. well-circumscribed where the nodule is located centrally in the lung without being connected to vasculature;
ii. vascularized where the nodule has significant connection(s) to the neighboring vessels while located centrally in the lung;
iii. pleural tail where the nodule is near the pleural surface, connected by a thin structure;
iv. and juxta-pleural where a significant portion of the nodule is connected to the pleural surface.

3.2 Nodule Simulation

In a CT scan the nodules can take various shapes and topologies, but the common characteristic amongst the nodules is the density distribution that tends to be concentrated around a region with an exponential decay (e.g., [7][15][16]). To illustrate this behavior, Fig. 3 shows the radial distance for each of the nodule types. Figure 4 shows the average distribution of HU densities for each nodule type. It can be observed that nodules' density distribution exhibit a bi-modal distribution (two dominant peaks) where the lower density mode represents regions surrounding the nodules and the higher density mode represents nodules densities. A linear model can be used to fit the joint density.

Using Gaussian kernels in the linear model, we employ the EM algorithm to obtain the joint and marginal density estimates. Figure 5 shows the marginal density estimates of the four nodule types. The range of intensity (HU) values is within two quantities q_{min} and q_{max} which will be used to estimate the density of a nodule template. From the knowledge of the nodule density (HU) vs. radial distance, and the nodules' marginal probabilities, we can compute the probability density function of the nodule's intensity (HU) with respect to radial distances from nodule centroid. These densities have been shown to be concentrated in radial distances around 5 pixels.

The behavior of the nodule intensity (HU) vs. radial distance provides three clues: 1) the intensity (HU) decays exponentially from the centroid; 2) the probability density of the intensity (HU) is concentrated within a range q_{min} and q_{max}; and 3) the probability density of the HU vs. radial density is concentrated below 10 pixels (hence, the spatial support of the templates is within 21x21 pixels). The first two clues

enable us to devise a formula for estimating the density of a nodule model (template). The third clue provides a figure of merit of how big the nodule template should be.

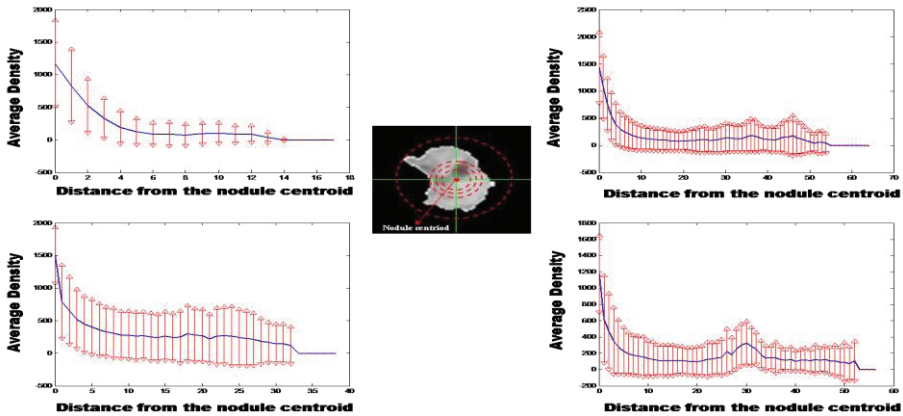


Fig. 3: *The distribution of the radial distance from the centroid of the nodules. The bars are one standard deviation off the mean values of each nodule type. Note the exponential behavior of the radial distance, and that it diminishes after a distance of 10.*

Given the range q_{min} and q_{max} of nodule density distribution (Fig. 5), obtained from the ground truth nodules (an ensemble of nodules outlined by the radiologists), the intensity or HU, at a distance *r* from the centroid, can be estimated by the following equations.

$$q(r) = q_{max} e^{-(r/\rho)^2}, \; 0 \leq r \leq \mathrm{R} \quad (1)$$

$$\rho = R(ln(q_{max}) - ln(q_{min}))^{-1/2} \quad (2)$$

These equations can be used to simulate parametric and non-parametric templates. For example, to simulate a circular template, we simply need to specify the radius R, and the intensity (HU) of the inside of the template at a distance *r* from the center can be calculated from Eq. 1 and 2. In this case, any of the four nodule intensity distributions in Fig. 5 may be used to obtain q_{min} and q_{max}. We should point out that this approach is valid for 2D and 3D non-parametric template designs as well.

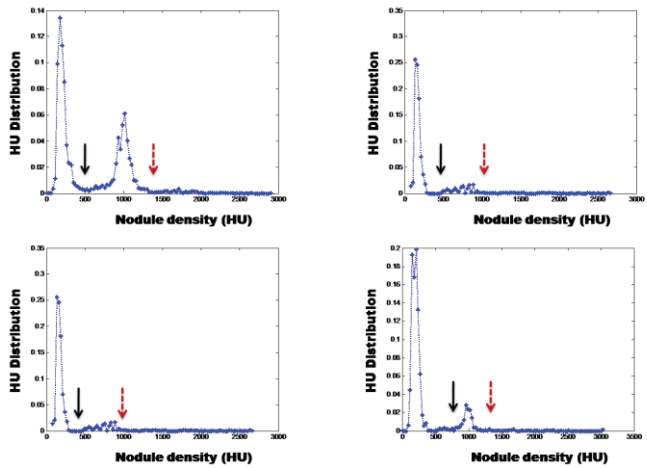


Fig. 4: *Average distribution of HU densities for each nodule type. The middle region corresponds to the nodule. The solid arrow is q_{min} and the dashed arrow (red) is q_{max} in Eq. 1 and 2.*

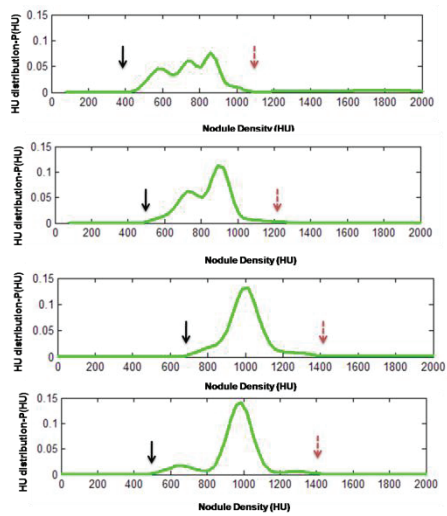


Fig. 5: *Probability density of the Radial distance of the nodules: Top left: Well-Circumscribed nodule; Top right: Vascular nodule; Bottom left: Juxta-Pleural; and Bottom right: Pleural-Tail nodule. Arrows show q_{min} and q_{max}. This behavior may be described by a single Gaussian (to resemble the circular-symmetric Gaussian in [7][15][16]).*

To test the overall approach of nodule detection, we used a set of 2D scans from the ELCAP clinical data set with known ground truth in terms of nodule type and location. Therefore, only 2D templates are used in this paper. The most common parametric nodule models in 2D are circular and semi-circular. Fig. 6 shows a few examples of such templates.

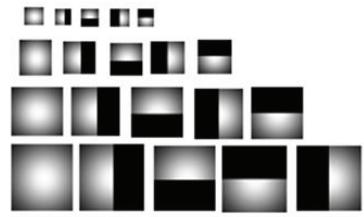


Fig. 6: *An ensemble of generated circular and semi-circular templates with radii ranging from 5 to 20 pixels with orientation from 0^o to 360^o with step size 90^o, in the semi-circular cases.*

The isotropic templates are defined in terms of the radius (size), *R*, and the gray level distribution as a circular symmetric Gaussian function [7][15][16] while non-isotropic templates are defined by radius, gray level distribution and orientation. Automatic generation of the gray level distribution of the nodules with known radius and histogram of nodule prototypes can be generated, for a given shape, using Eq. 1 and 2 above. This is particularly simple to perform in the case of parametric templates (e.g., Fig. 6, where given a radius *R* we only need to estimate the intensity of HU from these equations.

4. Template Matching

Among the widely used approaches for computerized nodule detection is the template matching approach, in which the nodule model (template) is swept across the scan (2D slices or the 3D volume) in a raster fashion and a similarity measure is calculated between the intensity (or HU) of the template and the region of the CT data underneath. If the result of the matching process between an unknown object in the CT data and the template is sufficiently high, the unknown object is labeled as resembling the template (i.e., a nodule candidate), however due to image noise, spatial, amplitude quantization effects, and a priori uncertainty of the exact shape and structure of the object to be detected, an exact match rarely occurs. Hence, a subsequent step of authenticating the detected nodule candidates is performed to reduce the false positives. Among the widely used similarity measures is the normalized cross correlation (NCC), which has a maximum value of unity that occurs if and only if the image function under the template exactly matches the template. The normalized cross-correlation of a template, *t(x,y)* with a sub-image *f(x,y)* is:

$$NCC = \frac{1}{n-1}\sum_{x,y}\frac{(f(x,y)-\bar{f})(t(x,y)-\bar{t})}{\sigma_f \sigma_t}, \qquad (3)$$

where *n* is the number of pixels in template *t(x,y)* and sub-image *f(x,y)* which are normalized by subtracting their means and dividing by their standard deviations.

The probability density functions (pdf) of nodule and non-nodule pixels are computed using the normalized cross correlation coefficients resulted from templates with varying parameters (shape, size and orientation if applicable). Based on the Bayesian

classification theory, the intersection between the pdf's of the two classes is chosen as the threshold separating the correlation coefficients resulted from nodule and non nodule pixels.

A decision fusion approach (Fig. 7) was also conducted in this paper using the circular and semi-circular templates in the detection process for radius 10 and 20 pixels with different orientations (semi-circular case) in a serial fashion and the final decision is Exclusive-OR (XOR) of the binary outputs. Template matching is performed as before using the same NCC threshold and the output of the template matching from each nodule model is a binary image (NCC values rank from zero to 1; after thresholding the zeros are NCC values below the specified threshold and the ones are otherwise), representing where the candidate nodules for the corresponding parametric template are located. Binarization is performed onto these images to give any pixels that are black a value of 0 and any white pixels a value of 1. The images are then Exclusively-ORed (XOR) together to receive a final black and white image that represents the candidate nodule locations for all of the nodule types. Fig. 8 summarizes this process.

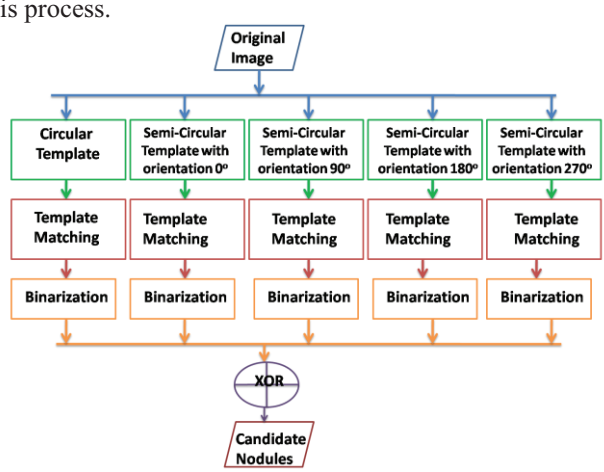


Fig. 7: *Block diagram of decision fusion template matching approach.*

XOR-ing the images after template matching and binarization provides us with locations where one of the template has a nodule detected for that pixel value will be taken as a true candidate nodule. We experimented with several other logic operators and studied the sensitivity and specificity results for each case before deciding on the XOR operator.

5 Experimental Results

This study is based on the ELCAP public database [10], which contains 50 sets of low-dose CT lung scans taken at a single breath-hold with slice thickness 1.25 mm. The locations of the 397 nodules are provided by the radiologists, where 39.12% are

juxta-pleural nodules, 13.95% are vascularized nodules, 31.29% are well-circumscribed nodules and 15.65% are pleural-tail nodules. The official reports indicate the mean nodule diameter to be 8.5 mm with standard deviation 3.6. The ELCAP database is of resolution 0.5x0.5mm [10].

The ELCAP expert identified a point on the nodule not the entire spatial support; this point may not necessarily coincide with the center or the centroid of the nodule. Indeed, the NCC measure decays as we move away from the spatial support of the nodules. The sensitivity and specificity of template matching was studied for templates of radii ranging from 0.5mm (1 pixel) to 10mm (20 pixels). The orientations of the templates in the semi-circular case are from 0 to 360 with step size of 45 degrees. For each detected nodule, let the coordinates of its centroid be $\boldsymbol{c}$ = (x,y) , the nodule is considered correctly detected and counted as true positive (TP) when the distance between the detected point and the closest ground truth point is smaller than the template radius. All other detected points are considered false positives (FP). True negative (TN) is the number of points which are not detected as candidate nodules and when compared to the ground truth they are not nodules. False negative (FN) is when no point is detected in the neighborhood of the ground truth nodule. The sensitivity and specificity are defined in terms of the false positive (FP) and the true positive (TP) nodules.

$$\text{Sensitivity } SN = TP/(TP+FN), \qquad \text{specificity } SP = TN/(FP+TN) \qquad (4)$$

These values are defined with respect to the NCC matching criterion. Fig. 8 shows the NCC distribution for parametric templates of radius 10 pixels (i.e., template size 21x21). As expected, higher values of NCC leads to reduced detection rates, while smaller values increases the detection rates, and consequently the number of FPs. A threshold of NCC value was set at 0.5 which have been shown, empirically, to be a good compromise between TP and FPs. This will be considered a hard classifier. The receiver operating characteristic (ROC) curve is used to plot the fraction of true positives (TPR = true positive rate) vs. the fraction of false positives (FPR = false positive rate), where TPR = TP = (TP+FN) and FPR = FP = (FP+TN). We experimented with different template sizes ranging from 1 to 20 pixels (0.5mm-5mm). The ROC curves showed us that variations in the template shape and/or orientation have minimal affect on well-circumscribed nodule detection. The semi-circular template with orientation 180 degrees best simulates juxta-pleural as well as pleural-tail, while the circular template least represents these nodule types.

Experiments revealed that larger template radii yield higher sensitivity; while the performance tends to decrease for smaller template sizes, yet, the specificity in the larger template radii overall decrease. The rate of performance decay depends on template shape and nodule type, the pleural-tail nodules are the most sensitive nodule type to the template size, while the well-Circumscribed nodule showed the least sensitivity to template size. At radius of 10, we found the overall sensitivity using the decision fusion approach to be higher than the results obtained without using the decision fusion approach. In the individual nodule cases the templates that yielded relatively smaller marginal differences from the decision fusion approach in terms of

sensitivity rates was the semi-circular template with orientation 315° representing the Well-Circumscribed nodule type and the circular template representing the Juxta-Pleural nodule type. The overall specificity was similar in the circular template cases. Semi-circular templates with orientations 0° to 135° (without decision fusion) provided higher specificity but this also means there are more false positives.

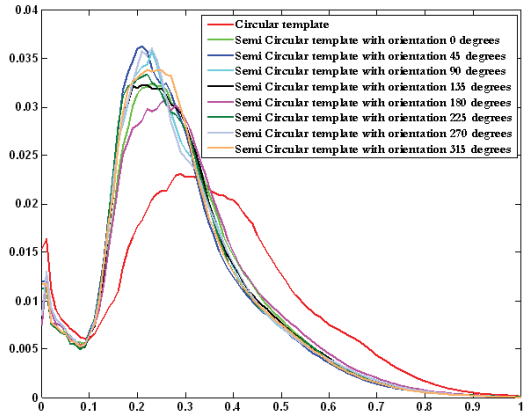


Fig. 8: *Distribution of the Normalized Cross-Correlation (NCC) for parametric nodules (circular and semi-circular) with radius 10 pixels. Higher NCC values results in less FPs while smaller values provide more FPs. At a threshold of 0.5 the range of NCC is {0.007, 0.012}.*

Tables below show the results for the sensitivity and specificity for radii 10 and 20 pixels. The sensitivity and specificity were analyzed for both the hard classifier and the decision fusion classifier. ROC curves are not included due to space limitations.

Nodule Type	Template Radius = 10 Hard Classifier		Template Radius = 10 Decision Fusion Classifier	
	Sensitivity	Specificity	Sensitivity	Specificity
Well-Circumscribed	49.44 %	81.15 %	58.43%	75.55%
Vascularized	70.73 %	78.54 %	73.17%	72.36%
Juxta-Pleural	83.48 %	66.73 %	93.04%	71.80%
Pleural-Tail	91.30 %	77.80 %	97.83%	72.81%

Nodule Type	Semicircular Template Radius 10; Orientation = 0° Hard Classifier		Semicircular Template Radius 10; Orientation = 90° Hard Classifier	
	Sensitivity	Specificity	Sensitivity	Specificity
Well-Circumscribed	37.08%	93.44%	31.46 %	93.70%
Vascularized	48.78 %	94.00 %	29.27 %	93.23%
Juxta-Pleural	75.65 %	88.41 %	73.04 %	86.70 %
Pleural-Tail	60.87%	92.50%	67.39%	92.50%

Nodule Type	Semicircular Template Radius 10; Orientation = 180° Hard Classifier		Semicircular Template Radius 10; Orientation =270° Hard Classifier	
	Sensitivity	Specificity	Sensitivity	Specificity
Well-Circumscribed	31.46 %	92.88 %	30.34%	93.64%

Vascularized	48.78 %	93.94 %	46.34%	93.36%
Juxta-Pleural	77.39%	86.30 %	75.65%	86.27%
Pleural-Tail	78.26 %	92.74 %	63.04%	92.80%

Nodule Type	Template Radius = 20 Hard Classifier		Template Radius = 20 Decision Fusion Classifier	
	Sensitivity	Specificity	Sensitivity	Specificity
Well-Circumscribed	76.40 %	61.17%	87.80%	36.58%
Vascularized	87.81%	53.71%	87.80%	36.58%
Juxta-Pleural	89.57 %	54.99%	92.17%	33.16%
Pleural-Tail	95.65 %	50.63%	97.83%	35.18%

Nodule Type	Semicircular Template Radius 20; Orientation = 0° Hard Classifier		Semicircular Template Radius 20; Orientation = 90° Hard Classifier	
	Sensitivity	Specificity	Sensitivity	Specificity
Well-Circumscribed	53.93 %	76.99%	57.30%	72.51%
Vascularized	75.61%	73.75%	73.17%	71.58%
Juxta-Pleural	82.61%	64.72%	80.87%	65.93%
Pleural-Tail	75.61%	73.75%	86.96%	66.32%

Nodule Type	Semicircular Template Radius 20; Orientation =180° Hard Classifier		Semicircular Template Radius 20; Orientation =270° Hard Classifier	
	Sensitivity	Specificity	Sensitivity	Specificity
Well-Circumscribed	59.55%	71.28%	53.93%	74.35%
Vascularized	70.73%	71.85%	73.17%	73.40 %
Juxta-Pleural	75.65%	64.24%	83.48%	66.30%
Pleural-Tail	86.96%	68.65%	91.30%	69.19%

6 Conclusion

In this paper, a data-driven approach was devised to model and simulate typical lung nodules. We studied the effect of template shape on detection of different nodules types. Variations in the template shape and/or orientation has minimal affect on well-circumscribed nodule detection. The hard classifier (based on thresholding the NCC) showed that the semi-circular template best simulates juxta-pleural and pleural-tail nodules, while the circular template least represents these nodule types. Similar conclusion also was achieved with the vascular nodules. Experiments revealed that larger template radii yield higher sensitivity, while the performance tends to decrease for smaller template sizes, yet, the specificity decreased with larger template. The overall performance depends on template shape and nodule type. The pleural-tail nodules are most sensitive to the template size, while the well-circumscribed nodule was the least sensitive. Overall, the decision fusion classifier provided best performance for templates of radius 10 pixels (i.e., nodule sizes of 1.5 mm), in terms of sensitivity and specificity. Current efforts will extend the nodule models into 3D using large clinical data set. Significant efforts are also directed towards designing non-parametric templates in 2D and 3D with attributes of the real data.

Acknowledgements: This research has been supported by the Kentucky Lung Cancer Program and The University of Louisville.

References

1. Kloecker, G., et al.: Lung Cancer in the US and in Kentucky. KMA **105** pp.159--164(2007)
2. Mountain, C. F.: Revisions in the international system for staging lung cancer. Chest **111** pp.1710--1717 (1997)
3. Gajra, A., et al.: Impact of tumor size on survival in stage IA non-small cell lung cancer: a case for subdividing stage IA disease. Lung Cancer **42** pp.51--57 (2003)
4. Zaho, B., Gamsu, G., Ginsberg M. S., Jiang, L., Schwartz, L. H.: Automatic Detection of small lung nodules on CT utilizing a local density maximum algorithm. Journal of Applied Clinical Medical Physics **4** (2003)
5. Armato, S. G. 3rd, Giger, M. L., Moran C. J., Blackburn, J. T., Doi, K., MacMahon H.: Computerized detection of pulmonary nodules on CT scans. Radio Graphics **19** pp.1303--1311 **(**1999)
6. S. Hu, E. A. Hoffman and J. M. Reinhardt, "Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images," IEEE Transactions on Medical Imaging, Vol. 20, pp. 490–498, 2001.
7. Y. Lee, T. Hara, H. Fujita, S. Itoh and T. Ishigaki, " Automated Detection of Pulmonary Nodules in Helical CT Images Based on an Improved Template-Matching Technique," IEEE Transactions on Medical Imaging, Vol. 20, 2001.
8. B. Ginneken, B. Romeny and M. Viergever, "Computer-Aided Diagnosis in Chest Radiography: A Survey," IEEE Transactions on Medical Imaging, Vol. 20, 2001.
9. I. Sluimer, A. Schilham, M. Prokop, and B. van Ginneken, "Computer Analysis of Computed Tomography Scans of the Lung: A Survey," IEEE Transactions on Medical Imaging, vol. 25, No. 4, pp. 385–405, April, 2006.
10. ELCAP public lung image database, http://www.via.cornell.edu/databases/lungdb.html
11. G. Grieg, O. Kubler, R. Kikinis, and F. A. Jolesz, "Nonlinear Anisotropic Filtering of MRI Data," IEEE Transactions on Medical Imaging, Vol. 11, No. 2, pp. 221-232, June 1992.
12. W. J. Kostis, et al., "Small pulmonary nodules: reproducibility of three-dimensional volumetric measurement and estimation of time to follow-up," Radiology, Vol. 231, pp. 446-52, 2004.
13. W. J. Kostis, A.P. Reeves, D. F. Yankelevitz and C. I. Henschke, "Three dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images," Medical Imaging IEEE Transactions Vol. 22, pp. 1259—1274, 2003.
14. Asem Ali and Aly A. Farag, "Automatic Lung Segmentation of Volumetric Low-Dose CT Scans Using Graph Cuts," 4th International Symposium on Visual Computing (ISVC-08), Las Vegas, December 1-3, 2008, pp. 258-267.
15. A. A. Farag, A. El-Baz, G.Gimelfarb, R. Falk and S. G. Hushek, "Automatic detection and recognition of lung abnormalities in helical CT images using deformable templates," Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI-2004, September 26-29, 2004, pp. 856-864.
16. Shireen Y. Elhabian, Amal A. Farag, Salwa Elshazly and Aly A Farag, "Sensitivity of Template Matching for Pulmonary Nodule Detection: A Case Study," (CIBEC 2008), Cairo, Egypt December 18-20, 2008, pp. 110 -114.

Measurement, Evaluation and Analysis of Wall Thickness of 3D Airway Trees across Bifurcations ⋆

Xiaomin Liu[1] ⋆⋆, Danny Z. Chen[1], Merryn Tawhai[2], Eric Hoffman[3], and Milan Sonka[3]

[1] Department of Computer Science & Engineering, University of Notre Dame, USA
xliu9@nd.edu
[2] Auckland Bioengineering Institute, The University of Auckland, New Zealand
[3] Iowa Institute for Biomedical Imaging, The University of Iowa, USA
milan-sonka@uiowa.edu

Abstract. Airway segmentation and analysis is an essential step in the understanding of pulmonary anatomy and early detection of lung diseases. In this paper, we evaluate a newly developed method for segmenting the inner and outer surfaces of the airway tree. This is the first approach reported to validate the segmentation of the double wall surfaces for the entire airway tree including bifurcations. Two approaches help analyze the tree topology: (1) 3D thinning algorithm which produces a one-voxel wide centerline, and (2) Delaunay triangulation based generation of both the inner and outer 3D medial axes (sheets). The two airway wall surfaces are identified using a 3D optimal graph search approach. Once segmented, the bifurcation/carina areas are defined based on (1) the surface point proximity to the branch points identified on the centerline, and (2) the surface point distance from the 3D outer medial sheets. The wall thickness measurements for both the bifurcation/ non-bifurcation and carina/non-carina points are grouped and reported as a function of tree generation. The accuracy of wall thickness measurement is assessed on CT-scanned double-wall physical phantoms. The airway wall thickness accuracy is similar for the bifurcation/carina and non-bifurcation/carina areas. The measurements on normal human *in vivo* data reflect reasonable and consistent differences of wall thickness from generation to generation.

1 Introduction

Measurements of airway dimensions such as the local diameters and airway wall thickness provide important information contributing to the understanding of lung pathology and early detection of lung diseases. For instance, airway wall

⋆ This research was supported in part by NSF grants CCF-0515203 and CCF-0916606 and NIH grants R01-EB004640 and R01-HL064368.
⋆⋆ The work of this author was supported in part by a graduate fellowship from the Center for Applied Mathematics, University of Notre Dame.

thickening is inversely related to the degree of airflow obstruction [1], and attenuates airway reactivity in patients with asthma [2]. However, while a number of segmentation and evaluation methods have been proposed for automatically segmenting and measuring the airway inner surface [3–5], few methods are known for segmentation of the airway outer surface and the measurement of wall thickness. This is especially true for measurement of wall thickness across airway tree bifurcations. In [6], the airway outer surface was segmented by locally deforming the lumen mesh under specific force constraints. However, the optimality of the solution is not guaranteed.

Li *et al.* [7] extended a graph search based approach of Wu and Chen [8] to segmenting multiple interrelated surfaces together, and successfully applied it to the segmentation of 3D airway double surfaces on the non-branching part. The solution of the graph search is optimal with respect to given cost functions. As a further extension, a new method was recently developed by Liu *et al.* [9] to identify both the inner and outer surfaces of the entire intrathoracic airway tree in 3D including the bifurcations. While the visual assessment of the inner/outer wall segmentation in *in-vivo* data suggested that the double surfaces are captured well by the method, quantitative validation of the segmentation accuracy remained a challenging problem. Manual tracing of airway tree wall surfaces is very hard to obtain in completeness, and even if attempted, suffers from low reproducibility and high inter- and intra-observer variability due to image ambiguity of the outer surface.

Because of the topological complexity, airway bifurcations are the most difficult area for accurate segmentation and measurement. It is also of clinical importance to improve assessment accuracy on these areas. For example, the experimentally observed local accumulations of particles within bronchial airway bifurcations may play a crucial role in lung cancer induction [10]. The complete bifurcation should be able to cover the entire "Y" shaped part that connects to the relatively straight branches (which are the non-bifurcation areas). In addition, it is physiologically and maybe diagnostically interesting to focus on specific parts of the bifurcation, such as the carina area between two sub-branches. For example, widening and distortion of the carina is a serious sign because it usually indicates carcinoma of the lymph nodes around the region where the trachea divides [11]. A good validation method should be able to distinguish the bifurcation/carina and non-bifurcation/carina areas in the airway trees, and evaluate the results separately in both these areas.

To cope with the above difficulties, we perform extensive validation using a CT-scanned double surface bifurcating phantom with known wall thickness to optimize the cost function of graph search [9]. Then the segmentation method in [9] (with the optimized cost functions) is applied to human *in vivo* CT data. The wall thickness is analyzed as a function of generation. In order to do so, we employ two methods to identify the bifurcation and carina areas on the airway tree based on the one voxel wide 3D skeleton and 3D medial axis sheets, respectively. Then, the wall thickness is measured on the identified bifurcation/non-bifurcation and carina/non-carina points, grouped and compared for different

generations. Consequently, meaningful information is produced about the variability and consistency of wall thickness across normal subjects for individual generations of the airway trees.

2 Segmentation

The multi-layer optimal graph search based algorithm is employed to segment both the inner and outer airway walls as an entire tree [9]. The method can be outlined as follows (see [9] for full details):

(1) Pre-segmentation. A pre-segmentation is needed to provide the basic information about the object's global structure. It is not necessary for the pre-segmentation to be locally accurate. However, it is crucial to preserve the topology of the target object. In our experiments, the airway trees are pre-segmented using commercially available Pulmonary Workstation PW+ software (VIDA Diagnostics, Oakdale, IA). Once a labeled image is generated by the pre-segmentation, it is transformed into a triangulated mesh using the marching cube algorithm.

(2) Image resampling. Based on the outcome of the pre-segmentation, the image is resampled on each vertex of the initial surface mesh, resulting in a set of vectors (called *columns*) of voxels. 3D medial axes (sheets) are computed based on Voronoi diagrams, which help to determine the directions and lengths of the resampling columns.

(3) Graph construction. Each voxel in the columns is considered as a node in the graph. There are three types of edges, representing the relations of voxels within the same surface or between different surfaces. A cost function is assigned to each node combining the first and second derivative. To distinguish properties of the inner and outer surfaces, two inversely oriented Sobel operators are applied to capture the directional information of the intensity changes.

(4) Graph search. A minimum s-t cut algorithm [8, 7] is applied to the resulting graph to simultaneously search for the inner and outer surfaces of airway walls simultaneously. This graph search identifies double wall surfaces simultaneously for all branches and all bifurcations of the entire tree in a single optimization process.

3 Identification of airway bifurcation

Starting from the trachea, the airway splits into left and right main bronchi, and continues to decrease size in ever-smaller branches at each bifurcation. The resulting tree structure consists of several generations. The segmentation around the bifurcation is especially difficult task because of the topological complexity. The advantage of the algorithm in [9] is the ability to segment the double surfaces around the bifurcation as described above. To evaluate the segmentation and wall thickness measurement for the bifurcation regions, the bifurcation areas

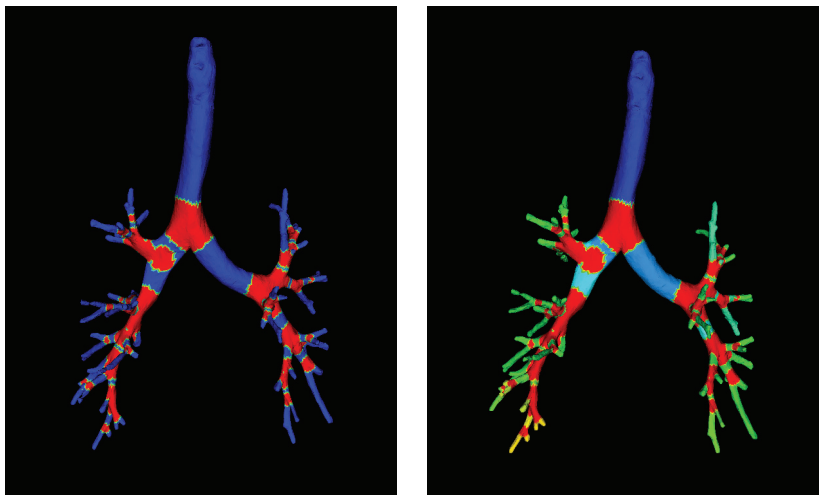
and other tree segments must be identified to allow performance comparisons in these two areas. 3D skeleton/medial axis is a powerful approach to provide a simplified structure of the object which would capture the same topology of the original object, and thus provide the information about the branching points, tree segment generation, etc.

3.1 3D thinning

Airway tree skeleton is obtained using a 3D thinning algorithm presented by Lee [12]. (See Fig. 1) The algorithm iteratively deletes the simple points while preserving the topology (connectivity, the same number of holes and cavities in the object). The output of the thinning algorithm is a one-voxel wide centerline. A tree-labeling process follows using the centerline that identifies the branching and non-branching points using a depth-first-search (DFS). Both branching and non-branching points are associated with a number representing the generation.

Fig. 1. 3D skeleton of an airway tree colored according to tree generations.

The bifurcations are identified by using spheres centered at each branching point. The sphere diameters are proportional to the local branch diameter. In Fig. 2, the diameter of the sphere is twice as large as the local diameter in order to include the entire bifurcation. In this situation, some of the branches in the 3^{rd}, 4-6^{th} generations that are short (comparing to the local diameter) may become mostly included in the bifurcation (covered in red).



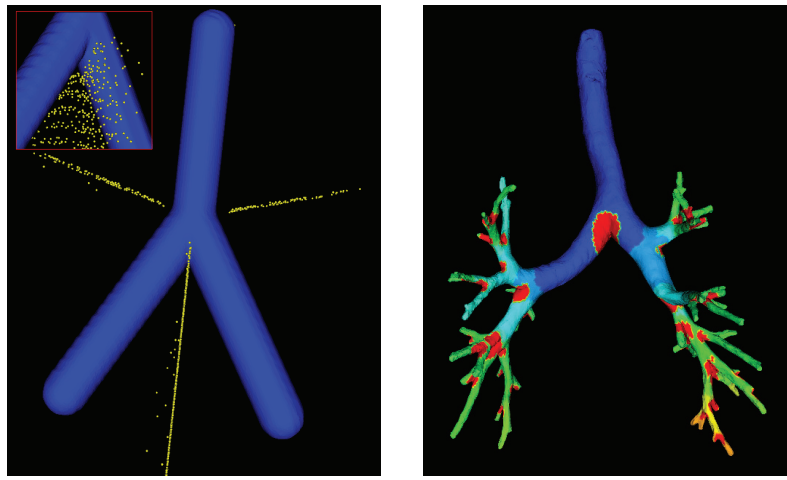
(a) (b)

Fig. 2. Airway trees with identified bifurcation and non-bifurcation areas. (a) Bifurcation areas are colored in red. (b) In this panel, bifurcation areas are colored in red while non-bifurcation areas are colored by generation.

As a well known phenomenon, the skeleton/medial axis is sensitive to local changes, which may cause undesirable fake branches in the skeleton. To get rid of these fake branches, we measured the length of each branch after the branching points are identified. Then, the branches that are short compared to local diameter are pruned out.

3.2 3D medial axis

3D thinning allows fast computation of the medial axis as it iteratively deletes all the surface points until a one-voxel wide centerline remains. Nevertheless, it lacks the ability to produce the outer medial axis, which may be important for capturing the changes in tree structure. While an approach based on Delaunay triangulation and Voronoi diagram of the surface may suffer from low computational efficiency, it produces both the inner and outer medial axis of the object at the same time and was chosen here.



(a) (b)

Fig. 3. Carina regions consist of points that are close to the outer medial axis. (a) Outer medial axis points (yellow) around a bifurcation – detail shown in the upper left corner. (b) Red areas represent carina regions. The tree is labeled with different colors for different generations.

The output of the graph search segmentation is a triangulated surface, where neighboring vertices/points are connected by edges. We can compute the Voronoi diagram and the dual Delaunay triangulation of S in 3D [13]. The medial axis is computed by using the poles in the Delaunay triangulation, which are selected from the centers of the big Delaunay balls adjacent to vertices. A pole is assigned to each of the mesh vertices by selecting the largest pole among the vertex k-nearest neighbors, in order to reduce the impact of possible noise on the surface [13].

Surface points that are close to the outer medial axis are selected. For an object with tree-like structure such as the airway, parental generations always split into two child airways at the site of bifurcation. The area where the two child airways separate forms a V-shape, which is the difficult region for wall thickness measurement because of local topological changes. In anatomy, the tracheal carina is a cartilage-rich saddle region within the trachea that separates the two mainstem bronchi. We define the carina regions in our measurements as the saddle areas where each generation branches into two child generations. As illustrated in Fig. 3, the carina regions can be well represented by points that are within a small distance from the outer medial axis (once such distance is determined properly by the local diameter). By selecting different cut-off values, the carina regions are not necessarily part of the bifurcation areas as defined in the last section.

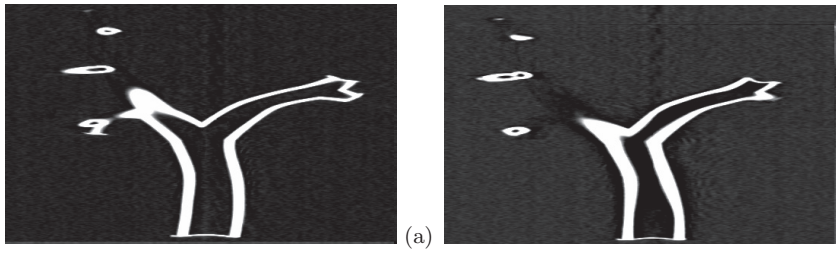


Fig. 4. Example 2D slices from two CT scans of the double-wall phantoms. (a) Phantom with smooth wall thickness distribution. (b) Phantom with non-smooth wall thickness distribution.

4 Experimental methods and results

4.1 Evaluation based on bifurcation and carina identification

After the labeling of the airway tree and identification of bifurcation and carina areas, we measure the wall thickness for the entire tree and group the results according to different generations in both bifurcation/non-bifurcation and carina/non-carina areas. The thickness is measured between the two segmented surfaces along the normal direction to the inner surface. The evaluation has been performed on two kinds of data: (1) two double-wall bifurcating phantoms derived from human *in vivo* data with smooth and non-smooth wall thickness, and (2) *in vivo* human CT scans.

Double wall phantoms The evaluation of wall thickness measurement is first performed on CT scans of two double wall phantoms with predefined wall thickness (Fig. 4). To construct the phantom, segmented volumetric images from multi-detector row computed tomography scanning of a healthy male lung are imported into the cmgui visualization software (www.cmiss.org). A bi-cubic Hermite finite element mesh is geometry-fitted to the surface of the outer wall. The mesh is given a volume by defining a wall thickness at each mesh node. The phantom is manufactured using a solid 3D printer (the Dimension Elite 3D Printer). The solid structure is composited of 0.254 mm thick layers of ABS plastic and support material. The support material is dissolved following manufacturing. The resolution of the discretized tetrahedral mesh is set at three levels as indicated in the results. The smooth and non-smooth phantoms have identical outer wall surfaces. The only difference in their geometry is the wall thickness at the mesh nodes and the derivatives for inner surface curvature. The inner surface derivatives in the non-smooth phantom are modified arbitrarily to create a challenging structure on which to test the algorithm. The wall thickness of both bifurcation and non-bifurcation areas is calculated and listed in the following table. The image size is $512 \times 512 \times 313$ and the voxel size is $0.23 \times 0.23 \times 0.7\text{mm}^3$.

The ground truth of the wall thickness is computed from the original digital mesh of the double wall phantom that is used for the fast prototyping of the

Table 1. Mean wall thickness errors for a double-surface bifurcating smooth phantom.

Smooth (mm)	Number of points	Unsigned error average(mm)	Signed error(mm)
Coarse	10374	0.27 ± 0.23	-0.023 ± 0.35
Original	40743	0.27 ± 0.23	-0.025 ± 0.35
Fine	73119	0.27 ± 0.23	-0.022 ± 0.35

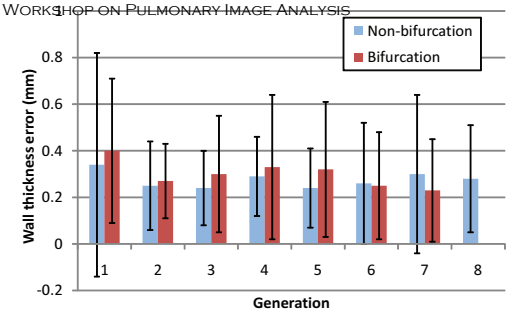
Table 2. Mean wall thickness errors for a double-surface bifurcating non-smooth phantom.

Nonsmooth (mm)	Number of points	Unsigned error average(mm)	Signed error(mm)
Coarse	11891	0.27 ± 0.22	-0.000 ± 0.35
Original	46537	0.27 ± 0.22	-0.001 ± 0.35
Fine	83800	0.27 ± 0.22	-0.004 ± 0.35

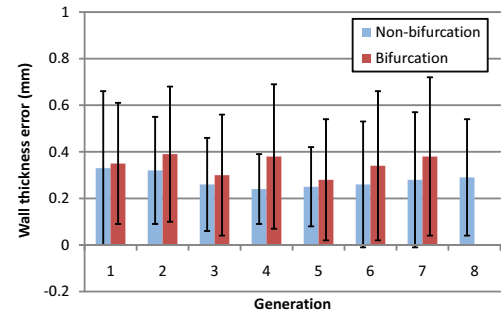
physical phantom. The mesh is then down-sampled to form a coarse mesh that is half as dense as the original one, and also up-sampled to a fine mesh twice as dense as the original one. As summarized in Tables 1 and 2, the mean errors and standard deviations of the wall thickness measurement are low (subvoxel) for both the smooth and non-smooth phantoms and thus do not depend on the mesh density. When considering the branching and non-branching areas on the phantom, the average errors and standard deviations of the wall thickness measurement are determined and shown in Fig. 5.

Human *in vivo* data The double-wall airway tree segmentation method is used to determine airway wall thickness for 7 sets of human CT scans from 7 normal subjects. The image size varies from $512 \times 512 \times 464$ to $512 \times 512 \times 587$, and the voxel size is $0.59 \times 0.59 \times 0.60\text{mm}^3$.

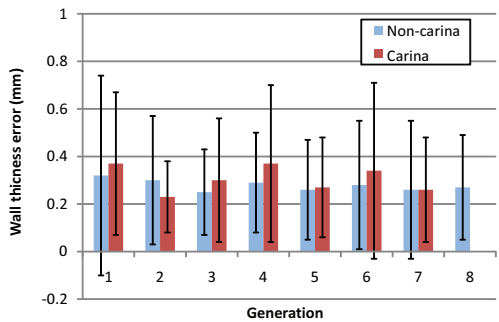
Fig. 6 (a) shows the average wall thickness at the bifurcation and non-bifurcation areas measured and grouped by airway tree generations. Fig. 6 (b) shows the wall thickness of the carina and non-carina areas, also as a function of generations. These figures depict a similar trend of the wall thickness, decreasing from the first to the sixth generation, that is consistent with the generational decrease of the local diameter. For the seventh to eleventh generations where the airway wall is relatively thin, wall thickness seems to stop decreasing. Comparing the measurements in the bifurcation and carina areas, we can see that the wall thickness of the bifurcation is larger than the wall thickness of the carina for the first six generations. The average wall thickness of the carina area is also smaller than non-carina area for the first a few generations. The measurements in the bifurcation and carina areas are very similar after the sixth generation.



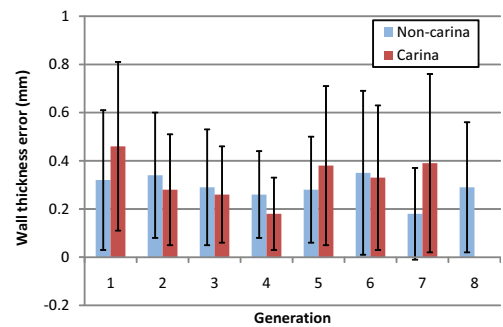
(a)



(b)



(c)



(d)

Fig. 5. The wall thickness measurement for the double-wall phantoms. The average errors and standard deviations are calculated for: (a) Bifurcation/non-bifurcation areas of the smooth phantom; (b) Bifurcation/non-bifurcation area of the non-smooth phantom; (c) Carina/non-carina areas of the smooth phantom; (d) Carina/non-carina areas of the non-smooth phantom.

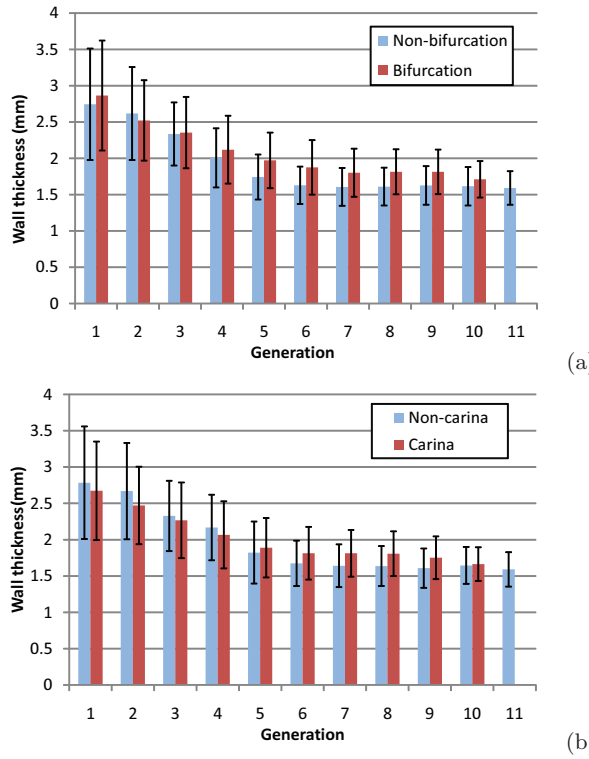


Fig. 6. Wall thickness as a function of airway tree generation. (a) Average wall thickness in bifurcation/non-bifurcation areas. (b) Average wall thickness in carina/non-carina areas. Error bars depict standard deviations.

5 Discussion and conclusion

An evaluation method of the double-surface airway tree segmentation is presented that simultaneously determines inner and outer surfaces of all branches and bifurcations for the entire tree. The surface points have been labeled with different generation automatically according to the skeleton. The complete bifurcation and relatively smaller carina area have been defined using two different methods of extracting the 3D medial axes/skeleton of the airway tree. By applying graph search based segmentation approaches, the experimental results on both the bifurcation/non-bifurcation and carina/non-carina areas achieve subvoxel accuracy in physical double-wall phantoms. The wall thickness analysis in the human *in vivo* data shows consistent decrease of mean wall thickness and standard deviation from parent to child generations. The proposed evaluation method is the first to allow separate measurement of airway wall thickness in the bifurcation and carina regions of the airway trees using a globally optimal

approach. Therefore, the reported method provides a new way for analyzing the airway wall properties in bifurcating regions. The quantification of wall properties in bifurcations allows novel disease-specific studies of intrathoracic airway tree physiology and function.

References

1. Deveci, F., Murat, A., Turgut, T., Altunta, E., Muz, M.H.: Airway wall thickness in patients with COPD and healthy current smokers and healthy non-smokers: assessment with high resolution computed tomographic scanning. Respiration **71** (2004) 602–610
2. Niimi, A., Matsumoto, H., Takemura, M., Ueda, T., Chin, K., Mishima, M.: Relationship of airway wall thickness to airway sensitivity and airway reactivity in asthma. American Journal of Respiratory and Critical Care Medicine **168** (2003) 983–988
3. Aykac, D., Hoffman, E.A., McLennan, G., Reinhardt, J.M.: Segmentation and analysis of the human airway tree from three-dimensional X-ray CT images. IEEE Transaction On Medical Imaging **22** (2003) 940–950
4. Fetita, C.I., Preteux, F., Beigelman-Aubry, C., Grenier, P.: Pulmonary airways: 3-D reconstruction from multislice CT and clinical investigation. IEEE Transaction on Medical Imaging **23** (2004) 1353–1364
5. Tschirren, J., Hoffman, E.A., McLennan, G., Sonka, M.: Intrathoracic airway trees: Segmentation and airway morphology analysis from low-dose CT scans. IEEE Transaction on Medical Imaging **24** (2005) 1529–1539
6. Saragaglia, A., Fetita, C., Brillet, P.Y., Preteux, F., Grenier, P.A.: Airway wall thickness assessment: A new functionality in virtual bronchoscopy investigation. In Manduca, A., Hu, X.P., eds.: Proc. of SPIE Medical Imaging 2007: Physiology, Function, and Structure from Medical Images. Volume 6511. (2007) 65110P
7. Li, K., Wu, X., Chen, D.Z., Sonka, M.: Optimal surface segmentation in volumetric images - A graph-theoretic approach. IEEE Transactions on Pattern Analysis and Machine Intelligence **28** (2006) 119–134
8. Wu, X., Chen, D.Z.: Optimal net surface problems with applications. In: Widmayer, P., Triguero, F., Morales, R., Hennessy, M., Eidenbenz, S., Conejo, R. (eds.) ICALP 2002. LNCS. (2002) 1029–1042
9. Liu, X., Chen, D.Z., Wu, X., Sonka, M.: Optimal graph search based image segmentation for objects with complex topologies. In Pluim, J.P., Dawant, B.M., eds.: Proc. of SPIE Medical Imaging 2009: Image Processing. Volume 7259. (2009) 725915
10. Balashazy, I., Hofmann, W., Heistracher, T.: Local particle deposition patterns may play a key role in the development of lung cancer. Journal of Applied Physiology **94** (2003) 1719–1725
11. Ellis, H.: Clinical anatomy: A revision and applied anatomy for clinical students. tenth edn. Blackwell (2002)
12. Lee, T.C., Kashyap, R.L., Chu, C.N.: Building skeleton models via 3-D medial surface/axis thinning algorithms. Computer Vision, Graphics, and Image Processing **56** (1994) 462–478
13. Dey, T.K., Sun, J.: Normal and feature estimations from noisy point clouds. Technical Report OSU-CISRC-7/50-TR50 (2005)

EXACT'09 Challenge Papers

Extraction of Airways from CT (EXACT'09)

Pechin Lo[1], Bram van Ginneken[2],
Joseph M. Reinhardt[3], and Marleen de Bruijne[1,4]

[1] Image Group, Department of Computer Science, University of Copenhagen, Denmark, pechin@diku.dk,
[2] Image Sciences Institute, University Medical Center Utrecht, The Netherlands,
[3] Department of Biomedical Engineering, Iowa Institute for Biomedical Imaging, The University of Iowa, USA,
[4] Biomedical Imaging Group Rotterdam, Departments of Radiology & Medical Informatics, Erasmus MC, Rotterdam, The Netherlands

Abstract. This paper describes a framework for evaluating airway extraction algorithms in a standardized manner and establishing reference segmentations that can be used for future algorithm development. Because of the sheer difficulty of constructing a complete reference standard manually, we propose to construct a reference using results from the algorithms being compared, by splitting each airway tree segmentation result into individual branch segments that are subsequently visually inspected by trained observers. Using the so constructed reference, a total of seven performance measures covering different aspects of segmentation quality are computed. We evaluated 15 airway tree extraction algorithms from different research groups on a diverse set of 20 chest CT scans from subjects ranging from healthy volunteers to patients with severe lung disease, who were scanned at different sites, with several different CT scanner models, and using a variety of scanning protocols and reconstruction parameters.

1 Introduction

Analysis of the airways in volumetric computed tomography (CT) scans plays an important role in the diagnosis and monitoring of lung diseases. While airway segmentation is a key component for such analysis, the accuracy and reliability of existing algorithms for segmenting the airway are still unknown because of the lack of a common test database, a reference, and standardized means for comparison.

The aim of this paper is to develop a framework for evaluating airway extraction algorithms in a standardized manner and to establish a database with reference segmentations that can be used for future algorithm development. Because of the sheer difficulty of establishing a complete reference standard manually, we propose to instead construct a reference using results from the algorithms to compare. Airway tree segmentations were first subdivided into their individual branches, which can be easily visualized and were subsequently evaluated by

trained observers. This way, manual annotation is avoided and the trained observer is given the easier task of deciding whether a segmented branch is correct or wrong, thus separating wrongly labeled regions from correctly labeled regions. We focus the evaluation on extracting the most complete airway trees, progressing into higher generation airways and extracting as many branches as possible. The exact airway shape and dimensions were not taken into account; branches were considered to be correct as long as there was no significant leakage outside the airway walls. Since several annotations of the same images were evaluated, the evaluation process can be accelerated by automatically accepting branches that have high overlap with previously accepted branches. Finally, a reference segmentation was established by taking the union of all correct branches.

A diverse set of chest CT scans was contributed by eight different institutions. A total of 40 scans were selected, with 20 of these scans designated as training set, reserved for algorithm training and/or parameter tuning, and the remaining 20 as test set. The images were selected to include a large variety of acquisition conditions and pathologies.

The comparative study was organized as a challenge at the 2nd International Workshop on Pulmonary Image Analysis, which was held in conjunction with MICCAI 2009. Invitations to participate were sent out to several mailing lists and to authors of published papers on airway segmentation. Due to the time required for manual evaluation, participants were asked to submit results from a single method only. A total of 22 teams registered to download the data, and 15 teams [1–15] submitted their results, ten in the fully automated category and five in the semi-automated category. All results submitted by participating teams were used to establish the reference used for evaluation.

2 Data

A total of 75 chest CT scans were contributed by eight different sites. Scans were taken on several different CT scanner models, using a variety of scanning protocols and reconstruction parameters. The condition of the scanned subjects varied widely, ranging from healthy volunteers to patients showing severe pathology of the airways or lung parenchyma. Among the contributed CT scans, 40 were selected to be included into the dataset for the challenge, which were further divided into a training set and a test set. Care was taken to ensure that scans of all eight sites were represented in both the training and test sets, all files were anonymized properly, no scans of the same subject would be in both the training and test sets, and both sets included the same number of scans of similar quality, obtained at the same site and with similar characteristics. The 20 images in the training set were named CASE01 to CASE20, and the 20 images in the test set were named CASE21 to CASE40. Table 1 presents some characteristics of the 20 test cases.

Table 1. Acquisition parameters of the 20 test cases. Dosage is presented as x-ray tube current (kVp) and exposure (mAs) pair. The breathing status indicates full inspiration (Insp.) or full expiration (Exp.). Contrast indicates whether intravenous contrast was used during acquisition. The abbreviations for the scanner models are as follows: Siemens Sensation (SS), Siemens Volume Zoom (SVZ), Philips Mx8000 IDT (PMI), Philips Brilliance (PB), Toshiba Aquilion (TA) and GE LightSpeed (GEL). * indicates that the scan is from the same subject as the previous scan.

	Thickness (mm)	Scanner	Convolution Kernel	Dosage	Breath state	Contrast
CASE21	0.6	SS64	B50f	200/100	Exp.	No
CASE22*	0.6	SS64	B50f	200/100	Insp.	No
CASE23	0.75	SS64	B50f	200/100	Insp.	No
CASE24	1	TA	FC12	10/5	Insp.	No
CASE25*	1	TA	FC10	150/75	Insp.	No
CASE26	1	TA	FC12	10/5	Insp.	No
CASE27*	1	TA	FC10	150/75	Insp.	No
CASE28	1.25	SVZ	B30f	300/100	Insp.	Yes
CASE29*	1.25	SVZ	B50f	300/100	Insp.	Yes
CASE30	1	PMI16	D	120/40	Insp.	No
CASE31	1	PMI16	D	120/40	Insp.	No
CASE32	1	PMI16	D	120/40	Insp.	No
CASE33	1	SS16	B60f	321/200	Insp.	No
CASE34	1	SS16	B60f	321/200	Insp.	No
CASE35	0.625	GEL16	Std.	441/6209	Insp.	No
CASE36	1	PB16P	C	206/130	Insp.	No
CASE37	1	PB16P	B	64/30	Insp.	No
CASE38*	1	PB16P	C	51/20	Exp.	No
CASE39	1	SS16	B70f	436/205	Insp.	Yes
CASE40	1	SS16	B70s	162/105	Insp.	No

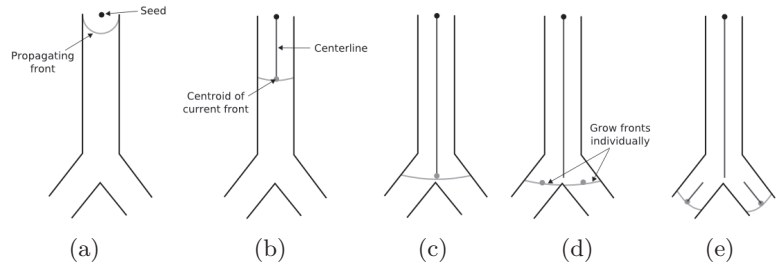


Fig. 1. Illustration of the process of dividing the airway tree into branches. (a) A seed point is set at the root of a tree and the front propagation process starts. (b) Centroid of the propagating front is stored as centerline at each time step. (c) The centerline is stored up to the point before a bifurcation. (d) Bifurcation is detected as the front splits. The individual fronts are used as seeds to perform front propagation in each of the branches. (e) Front propagation proceeds in each branch.

3 Airway branch evaluation

3.1 Subdividing an airway tree into branches

An important component of our evaluation is the process of subdividing an airway tree into its individual branch segments. This was done by detecting the bifurcations using wave front propagation, similar to [16]. The key idea is that a wave front propagating through a tree structure remains connected until it encounters a bifurcation, and any side branches can thus be detected as new disconnected components in the propagating front.

The front was propagated using a fast marching algorithm [17], with the speed function set to 1 within the segmented structure and 0 outside. We monitor the front through a set of "trial" points in the fast marching process. Connected component analysis was applied to the trial points when the time stamp from the fast marching algorithm increased by $1/\Delta D$, where ΔD is the distance between two voxels. Propagation stops when multiple disconnected components were detected in the front, whereupon the process was repeated on the individual split fronts to obtain the branches at the next level. The process ends when all marked regions of the tree have been evaluated. During the front propagation, the centerlines were obtained by storing the centroid of the front at every step. Figure 1 illustrates the algorithm.

3.2 Visual assessment

To enable visual inspection of extracted branches, each of the branches was presented to the trained observers using a fixed number of slices through the branch at different positions and orientations. Two different views were used, a reformatted and a reoriented view. The reformatted view was obtained by straightening the centerline of the branch segment, while the reoriented view

was obtained by rotating the branch segment such that its main axis coincides with the x-axis.

A total of eight slices were extracted from the reformatted view. The first four slices were taken perpendicular to the centerline, distributed evenly from the start to the end of the centerline. The other four slices were taken along the centerline, at cut planes that were angled at 0°, 45°, 90° and 135°. A schematic view of the cut planes for the different slices for the reformatted view is shown in Figure 2(b), with slice examples in Figure 2(d).

For the reoriented view, a total of nine slices were extracted, three from each axis. For the y-axis and z-axis, which are perpendicular to the axis of the branch segment, the cut planes were placed at 15%, 50% and 85% of the width of the branch measured in the respective axis. On the x-axis, the cut planes were placed at 5%, 50% and 95% of the length of the branch. Figure 2(b) shows the cut planes for the slices extracted from reoriented view, and examples of the slices are shown in Figure 2(e).

3.3 Evaluation of individual branches

Based on the slices from the two views described in Section 3.2, the trained observers were asked to assign one of the four following labels to each branch: "correct", "partly wrong", "wrong" or "unknown". A branch is "correct" if it does not have leakage outside the airway wall. "Partly wrong" is assigned to a branch if part of the branch lies well within the airway lumen, while part is outside the walls. A branch is "wrong" if it does not contain airway lumen at all. The "unknown" label is used when the observers are unable to determine whether a branch is an airway or not.

The evaluation of each branch was performed in two phases. At phase one, two observers are assigned to evaluate the branch. If both observers assign the same label, the evaluation is complete. Otherwise the evaluation proceeds to phase two, where three new observers are assigned. In this phase, the final label assigned to the branch is the label that constitutes the majority vote among the three new observers. In case of no majority, the branch is labeled as "unknown". The entire process was automated by web-based distribution of tasks to a pool of ten trained observers, who labeled the 15 segmentations of each of the 20 test cases. The observers were all medical students who were familiar with CT and chest anatomy. They were trained with a set of examples and subsequently processed two complete segmentations of different scans to ensure their ratings were reliable.

4 Establishing a reference

The reference segmentation for a CT scan was constructed by fusing branches based on the labels assigned as described in Section 3, which involves evaluated branches of segmented airway trees obtained from all participating teams.

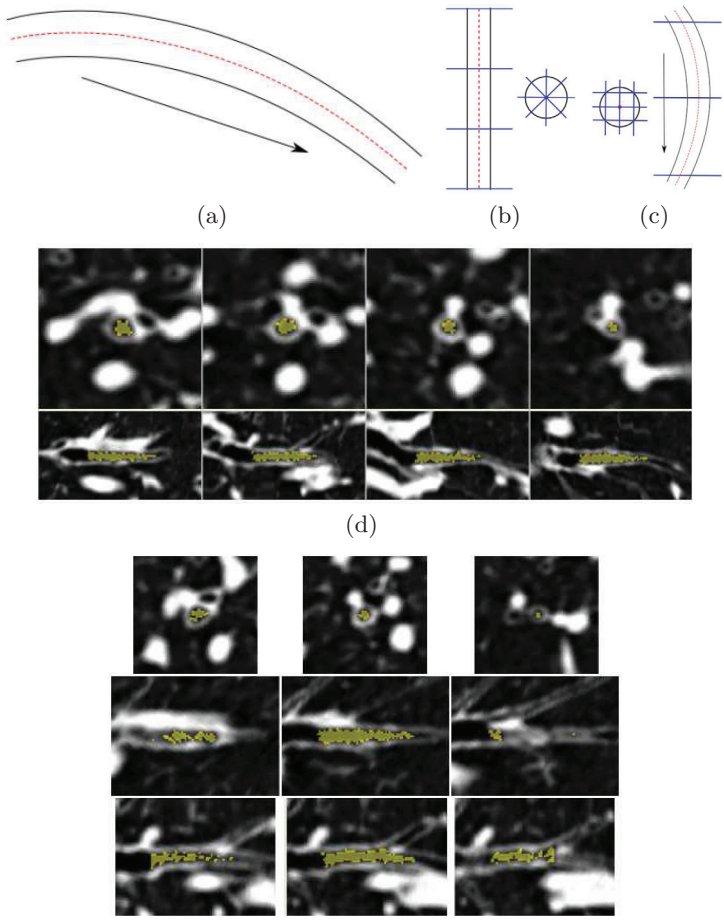


Fig. 2. Schematics showing the (a) original airway, (b) reformatted and (c) reoriented views. The arrow is the main orientation of the airway and the cut planes are shown in blue. Example of the slices extracted: (d) Reformatted view; top row are slices perpendicular to the branch segment and bottom row are slices taken in the direction of the branch segment. (e) Reoriented view; top row are axial slices, and middle and bottom row are cut planes parallel to the branch axis.

4.1 Automated branch acceptance

To speed up the evaluation process and reduce the number of branches that need to be evaluated by the observers, branches that were very similar to branches previously labeled "correct" were accepted automatically.

The following two criteria were used to determine whether to exempt a branch from manual evaluation or not:

1. Centerline overlap: Every point in the centerline is within a 26-neighborhood to a "correct" voxel.
2. Volume overlap: More than 80% of the voxels of the branch are labeled as "correct" in the reference.

Branches that fulfilled both criteria were labeled as "correct" and exempted from manual evaluation.

4.2 Updating the reference

The labels assigned by the observers as described in Section 3.3 and/or assigned automatically as above were subsequently stored in a reference image. The labels were stored as integer values with $L_c > L_w > L_p > L_u > 0$, where L_c, L_w, L_p and L_u represent the the values for "correct", "wrong", "partly wrong" and "unknown" respectively.

A reference image of a given CT scan started out as an empty image, with all voxels having value 0. Evaluated branches from a new segmentation of the same scan were added into the reference by applying the following update rule on all voxels $\boldsymbol{x}$

$$G_{t+1}(\boldsymbol{x}) = max\left(G_t(\boldsymbol{x}), L(\boldsymbol{x})\right),$$

where G_{t+1} and G_t are the reference after and before the update respectively, and L is the label assigned by the observers. The final reference of a CT scan was obtained once it was updated with all available evaluated segmentations.

4.3 Final reference

The final reference segmentation was defined as all voxels that were labeled as "correct" after updating the reference with all available evaluated segmentation results. Consequently, some of the voxels that were part of an airway lumen, but belonged to a branch segment containing errors and labeled "partly wrong", may be discarded. However, in our case, where a number of segmented airway trees from different algorithms are processed, the vast majority of airway voxels that were once labeled as "partly wrong" in an intermediate step are relabeled as they overlap with "correct" regions from another segmentation.

5 Evaluation

The algorithm described in Section 3.1 was used to obtain the centerlines of the branches in the reference segmentation. To enable comparison of the results obtained by different algorithms, we standardized the centerlines from the results to those of the reference segmentations. This was done by projecting the centerlines from a submitted result to the corresponding centerlines in the reference. The detected length of a given branch in the reference can then be obtained by measuring the length of the overlapping sections. To ensure that a branch is not trivially detected, a branch in the reference is said to be detected by a segmented airway tree only when the detected length is greater than $\Delta l = 1$ mm.

Seven performance measures were computed from the results of all participating teams:

1. Branch count: the number of branches that are detected correctly.
2. Branch detected: the fraction of branches that are detected, with respect to the total number of branches present in the reference.
3. Tree length: the sum of the length of the centerlines of all correctly detected branches.
4. Tree length detected: the fraction of tree length that is detected correctly relative to the reference.
5. Leakage count: the number of disconnected (26-connectivity) sources where leakage occurs. We define a leakage source as "correct" voxels that have at least a non-"correct" voxel within its 26-neighbors. A dilation of one voxel was applied to the detected sources prior to searching for the disconnected components in order to remove trivially disconnected sources.
6. Leakage volume: the volume of regions that are not marked as "correct" in the reference.
7. False positive rate: the fraction of the total segmented volume that is not marked as "correct" relative to the reference.

The trachea was excluded from measures $1-4$. For measures $5-7$, the trachea, left and right main bronchi were excluded.

6 Results

Table 2 presents the seven evaluation measures for the 15 participating teams, averaged over the 20 test cases. Figure 3(a) gives an overview of the average performance of the participating teams in terms of false positive rate over tree length detected, while Figure 3(b) shows the relation between the leakage volume and the number of leaks detected. Box plots of tree length detected and false positive rate for the different teams are given in Figure 4, while box plots for the number of branches detected and the leakage volume for each of the scans are given in Figure 5. In the box plots, the median is indicated by the red line, and the 25th and 75th percentile are indicated by the lower and upper edges of the box respectively. The lines below and above the box, or "whiskers", represent

the largest and smallest values that are within 1.5 times the interquartile range, while the red open circles show all outliers outside this range. Surface renderings from two cases are given in Figure 6 and 7, with correct and wrong regions indicated in green and red respectively.

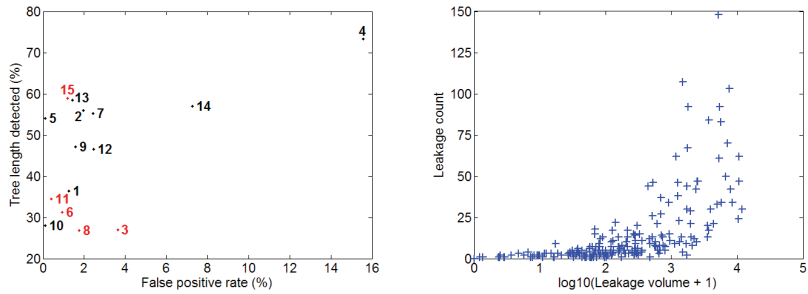
Table 2. Average evaluation measures from each team average over the twenty cases in the test set. * indicates semi-automated category and best result for each measurement are indicated in bold.

	Branch count	Branch detected (%)	Tree length (cm)	Tree length detected (%)	Leakage count	Leakage volume (mm^3)	False positive rate (%)
Team [1]	91.1	43.5	64.6	36.4	2.5	152.3	1.27
Team [2]	157.8	62.8	122.4	55.9	12.0	563.5	1.96
Team [3]*	74.2	32.1	51.9	26.9	4.2	430.4	3.63
Team [4]	**186.8**	**76.5**	**158.7**	**73.3**	35.5	5138.2	15.56
Team [5]	150.4	59.8	118.4	54.0	1.9	18.2	0.11
Team [6]*	77.5	36.7	54.4	31.3	2.3	116.3	0.92
Team [7]	146.8	57.9	125.2	55.2	6.5	576.6	2.44
Team [8]*	71.5	30.9	52.0	26.9	0.9	126.8	1.75
Team [9]	139.0	56.0	100.6	47.1	13.5	368.9	1.58
Team [10]	79.3	32.4	57.8	28.1	**0.4**	**14.3**	**0.11**
Team [11]*	93.5	41.7	65.7	34.5	1.9	39.2	0.41
Team [12]	130.1	53.8	95.8	46.6	5.6	559.0	2.47
Team [13]	152.1	63.0	122.4	58.4	5.0	372.4	1.44
Team [14]	161.4	67.2	115.4	57.0	44.1	1873.4	7.27
Team [15]*	148.7	63.1	119.2	58.9	10.4	158.8	1.19

7 Discussions

As expected, there appears to be a trade off between sensitivity and specificity in the airway tree extraction, as can be observed from Figure 3(a) and Figure 4, where more complete trees are usually accompanied by a larger percentage of false positives. The most conservative algorithm is that of team [10], which obtains the smallest average leakage volume (14.3 mm^3) with one of the shortest trees (57.8 cm), while the algorithm of team [4] is the most explorative one, yielding the longest trees (158.7 cm) but at the expense of the largest leakage volume (5138.2 mm^3).

In general, the semi-automatic methods perform no better than the fully automatic methods. The manual interaction in most semi-automatic methods is limited to selecting initial seed points for the trachea [11, 8] or manually adapting parameters [6, 3]. An exception is the algorithm of [15], which allows for extensive editing of results.



(a) (b)

Fig. 3. (a) Average false positive rate versus tree length detected of all participating teams, with teams in the semi-automated category in red. (b) Logarithm of leakage volume versus leakage count from all test cases results.

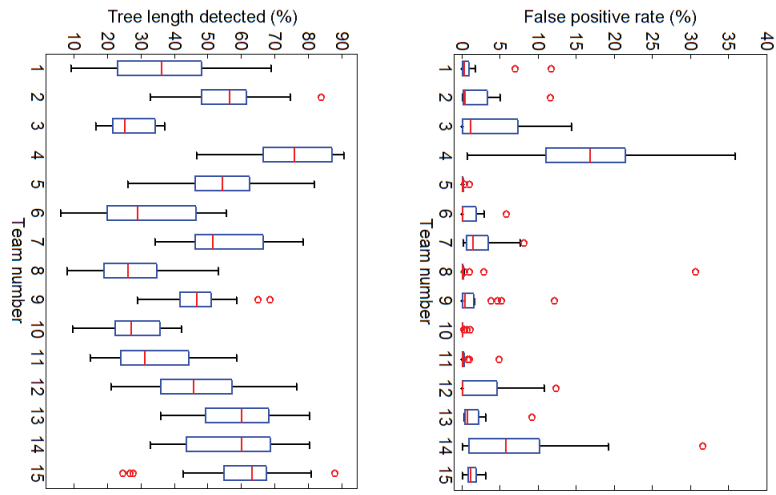
Figure 3(b) shows that the number of sources of leakage in an airway tree segmentation is often quite small. We found that over 66% of the submitted results have a leakage count of less than 5, with a maximum leakage volume of 1740.72 mm^3. This suggests that manually removing leakages from automatically extracted airway trees may be viable for most algorithms, provided that proper tools are available.

Interestingly, no algorithm came close to detecting the entire reference tree, as observed from Figure 4(a). The highest branch count and tree length detected for each case range from 64.6% to 94.3% and 62.6% to 90.4% respectively, with the average measures for each team no higher than 77%. This suggests that better results may be achieved by combining algorithms that use different approaches.

The box plots of the results per test case in Figure 5 reveal that in general, ultra low dose scans (cases 24, 26, 37, and 38) are more difficult to segment than low dose or clinical dose scans. From these scans, significantly fewer branches were extracted ($p < 0.01$ from Student's t-tests) with no significant difference in leakage volume ($p = 0.10$). For the two pairs of low dose and ultra low dose scans (case 24, 25 and 26, 27), branch count were significantly higher ($p < 0.01$ from paired Student's t-tests) for the low dose scans (93 branches) than for the ultra low dose (73 branches) scans, while there was no significant difference in leakage volume ($p = 0.54$).

From the available paired inspiration and expiration scans (case 21, 22 and 37, 38), segmentations of the inspiration scans include more correct branches but also more leakage than their expiration counterparts, where inspiration scans exhibits an average branch count of 145 branches and leakage volume of 942 mm^3 as compared to 76 branches and 115 mm^3 from expiration scans. A paired Student's t-tests shows that these difference are indeed significant, with $p < 0.01$ for both branch count and $p = 0.02$ for leakage volume.

One image pair, cases 28 and 29, is the same scan reconstructed using a soft and a hard kernel. For these scans, significantly more branches ($p < 0.01$)



(a) (b)

Fig. 4. Box plots of (a) tree length detected and (b) false positive rate of the teams.

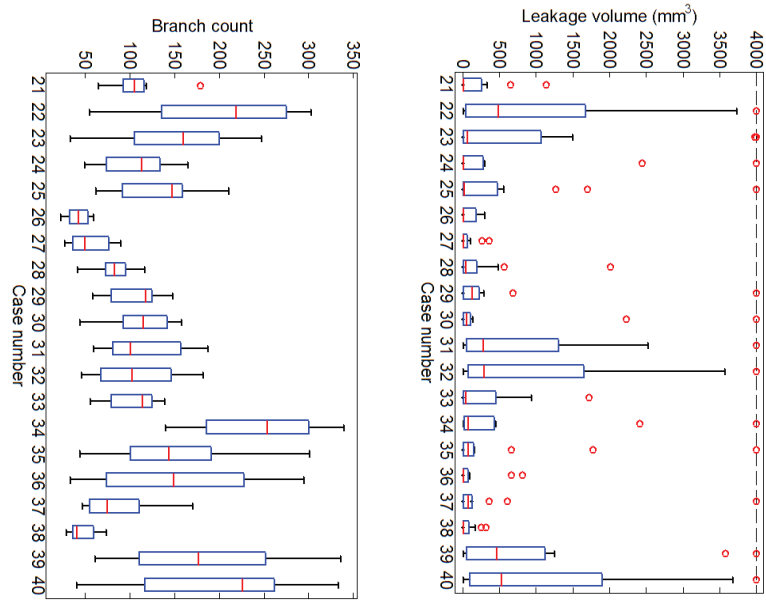
were extracted from the scan constructed using the hard kernel, average of 106 branches compared to 80 branches. Although average leakage volume for hard kernel scan was higher, 418 mm^3 compared to 236 mm^3, the difference was not significant ($p = 0.30$).

8 Conclusion

A framework for evaluating airway extraction algorithms in a standardized manner and establishing a reference segmentation is presented. Results obtained by 15 different airway tree extraction algorithms on a diverse set of 20 chest CT scans were manually evaluated and a reference was constructed from all correctly segmented branches. A total of seven performance measures were computed for each of the algorithms using the constructed reference. Results showed that no algorithm is capable of extracting more than 77% of the reference, in terms of both branch count and tree length, on average, indicating that better results may be achieved by combining results from different algorithms.

Acknowledgments. We would like to thank the following people for contributing data to this study:

– Haseem Ashraf and Asger Dirksen (Gentofte University Hospital, Denmark)
– Jesper Pedersen (Rigshospitalet - Copenhagen University Hospital, Denmark)



(a) (b)

Fig. 5. Box plots of (a) branch count and (b) leakage volume, with the maximum leakage volume clipped at 4000 mm^3, of the 20 test cases.

– Patrik Rogalla (Charité, Humboldt University Berlin, Germany)
– Jan-Martin Kuhnigk (Fraunhofer MEVIS, Germany)
– Rafael Wiemker, Cristian Lorenz (Philips Research Lab Hamburg, Germany)
– Berthold Wein (University Hospital of Aachen, Germany)
– Ken Mori (Graduate School of Information Science Nagoya University, Japan)
– Ieneke Hartmann (Erasmus MC - University Medical Center Rotterdam, The Netherlands)
– Eric Hoffman (Department of Radiology, University of Iowa, USA)
– Atilla Kiraly, Carol Novak, Benjamin Odry (Siemens Corporate Research, USA)
– David Naidich (NYU Medical Center, USA)
– Mathias Prokop (University Medical Center Utrecht, the Netherlands)

We also thank all participants of EXACT'09 and the ten trained observers for making this evaluation possible.

This work was funded in part by the Danish Council for Strategic Research (NABIIT), the Netherlands Organization for Scientific Research (NWO), and by grants HL080285 and HL079406 from the U.S. National Institutes of Health.

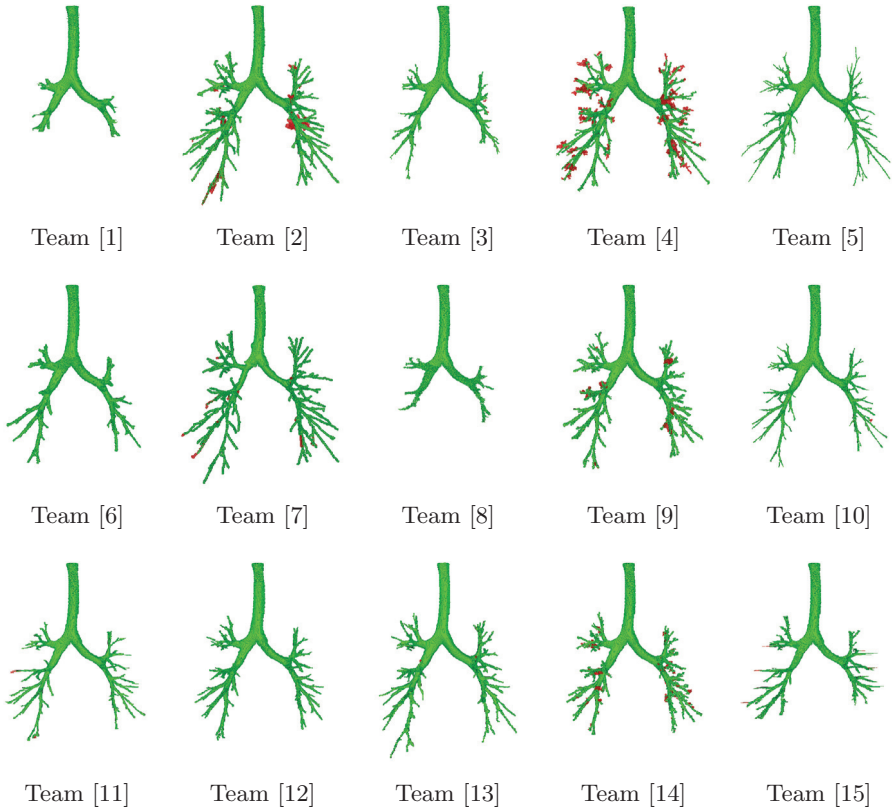


Fig. 6. Surface renderings of results for case 23, with correct and wrong regions shown in green and red respectively.

References

1. Irving, B., Taylor, P., Todd-Pokropek, A.: 3D segmentation of the airway tree using a morphology based method. In: Proc. of Second International Workshop on Pulmonary Image Analysis. (2009)
2. Fetita, C., Ortner, M., Brillet, P.Y., Prêteux, F., Grenier, P.: A morphological-aggregative approach for 3D segmentation of pulmonary airways from generic MSCT acquisitions. In: Proc. of Second International Workshop on Pulmonary Image Analysis. (2009)
3. Pinho, R., Luyckx, S., Sijbers, J.: Robust region growing based intrathoracic airway tree segmentation. In: Proc. of Second International Workshop on Pulmonary Image Analysis. (2009)
4. Feuerstein, M., Kitasaka, T., Mori, K.: Adaptive branch tracing and image sharpening for airway tree extraction in 3-D chest CT. In: Proc. of Second International Workshop on Pulmonary Image Analysis. (2009)
5. Lo, P., Sporring, J., de Bruijne, M.: Multiscale vessel-guided airway tree segmentation. In: Proc. of Second International Workshop on Pulmonary Image Analysis. (2009)

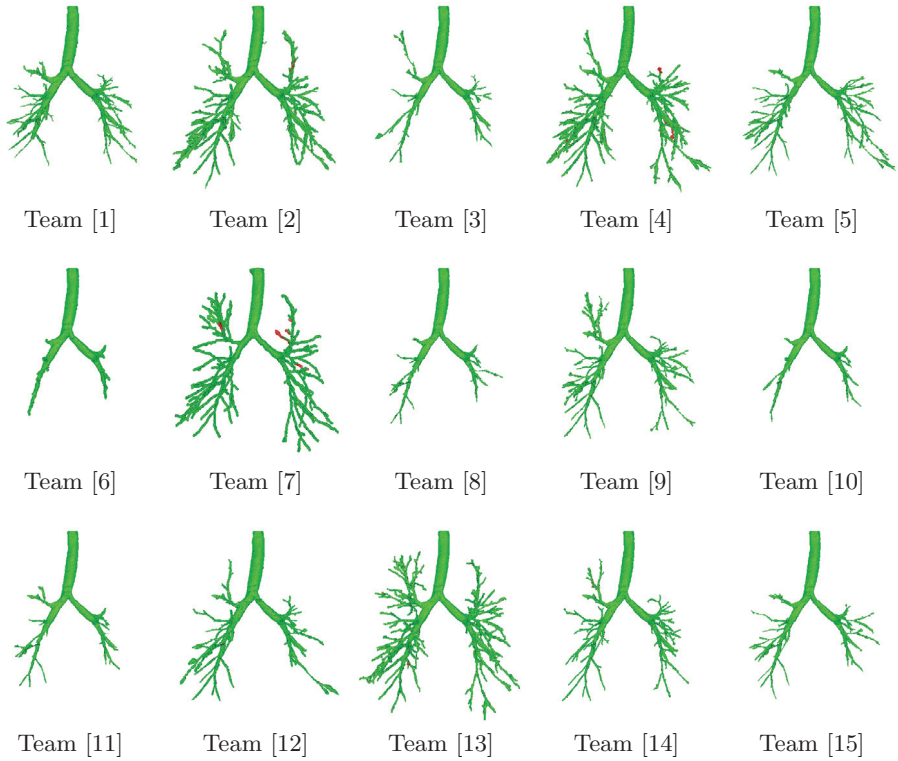


Fig. 7. Surface renderings of results for case 36, with correct and wrong regions shown in green and red respectively.

6. Fabijanska, A.: Results of applying two-pass region growing algorithm for airway tree segmentation to MDCT chest scans from EXACT database. In: Proc. of Second International Workshop on Pulmonary Image Analysis. (2009)
7. Bauer, C., Pock, T., Bischof, H., Beichel, R.: Airway tree reconstruction based on tube detection. In: Proc. of Second International Workshop on Pulmonary Image Analysis. (2009)
8. Mendoza, C.S., Acha, B., Serrano, C.: Maximal contrast adaptive region growing for CT airway tree segmentation. In: Proc. of Second International Workshop on Pulmonary Image Analysis. (2009)
9. Wiemker, R., Buelow, T., Lorenz, C.: A simple centricity-based region growing algorithm for the extraction of airways. In: Proc. of Second International Workshop on Pulmonary Image Analysis. (2009)
10. Lee, J., Reeves, A.P.: Segmentation of the airway tree from chest CT using local volume of interest. In: Proc. of Second International Workshop on Pulmonary Image Analysis. (2009)
11. Born, S., Iwamaru, D., Pfeifle, M., Bartz, D.: Three-step segmentation of the lower airways with advanced leakage-control. In: Proc. of Second International Workshop on Pulmonary Image Analysis. (2009)

12. Weinheimer, O., Achenbach, T., Düber, C.: Fully automated extraction of airways from CT scans based on self-adapting region growing. In: Proc. of Second International Workshop on Pulmonary Image Analysis. (2009)
13. Bauer, C., Bischof, H., Beichel, R.: Segmentation of airways based on gradient vector flow. In: Proc. of Second International Workshop on Pulmonary Image Analysis. (2009)
14. van Rikxoort, E.M., Baggerman, W., van Ginneken, B.: Automatic segmentation of the airway tree from thoracic CT scans using a multi-threshold approach. In: Proc. of Second International Workshop on Pulmonary Image Analysis. (2009)
15. Tschirren, J., Yavarna, T., Reinhardt, J.: Airway segmentation framework for clinical environments. In: Proc. of Second International Workshop on Pulmonary Image Analysis. (2009)
16. Schlathölter, T., Lorenz, C., Carlsen, I.C., Renisch, S., Deschamps, T.: Simultaneous segmentation and tree reconstruction of the airways for virtual bronchoscopy. Volume 4684., SPIE (2002) 103–113
17. Malladi, R., Sethian, J.: Level set and fast marching methods in image processing and computer vision. In: Proc. International Conference on Image Processing. Volume 1. (1996) 489–492 vol.1

Segmentation of Airways Based on Gradient Vector Flow

Christian Bauer[1,2], Horst Bischof[1], and Reinhard Beichel[2,3,4]

[1] Inst. for Computer Graphics and Vision, Graz University of Technology, Austria
[2] Dept. of Electrical and Computer Engineering
[3] Dept. of Internal Medicine
[4] The Iowa Institute for Biomedical Imaging
The University of Iowa, Iowa City, IA 52242, USA
{cbauer, bischof}@icg.tu-graz.ac.at
reinhard-beichel@uiowa.edu

Abstract. We present an automated approach for the segmentation of airways in CT datasets. The approach utilizes the Gradient Vector Flow and consists of two main processing steps. Initially, airway-like structures are identified and their centerlines are extracted. These centerlines are used in a second step to initialize the actual segmentation of the corresponding airways. An evaluation on 20 clinical datasets shows that our method achieves a good average airway branch count (63.0%) without any major leakage.

1 Introduction

Segmentation of airway trees in CT is of importance for various clinical applications, and several methods have been presented in the literature for this task. An overview can be found in the survey of Sluimer et al. [1]. A summary of newer approaches can be found in [2]. While some methods focus on accurate segmentation of airways, others are primarily targeting the extraction of airway tree skeletons that are of importance for applications like virtual bronchoscopy [3] or airway tree labeling and anatomical matching [4].

In our previous works [5, 6], we presented a generic framework for the direct extraction of complete curve skeletons of branched tubular structures from gray-value images. The approach utilizes the Gradient Vector Flow (GVF) [7] – an anisotropic edge preserving gradient diffusion method – to detect tubular objects [5] and to extract their associated medial curves [6]. It does not require a prior segmentation and extracts curve skeletons of comparable quality to sophisticated skeletonization methods applied to segmentations.

The GVF – that represents a core component of the approach – was originally presented to guide snake based segmentations [7]. In the literature, properties of the GVF (or similar gradient diffusion methods) have also been used differently, for example to generate voxel accurate 2D and 3D segmentations without using snakes [8–11], and issues related to initialization of the segmentation or deficiencies of the GVF in case of boundary concavities, as they would occur with

side branches in branched tubular networks, have been addressed. For example, Li et al. [11] used a gradient flow tracking in the GVF field in combination with a locally adaptive thresholding scheme based on gray value statistics to segment blob like 3D structures. To our knowledge, none of the so far presented methods is directly applicable for segmentation of tubular structures like airways.

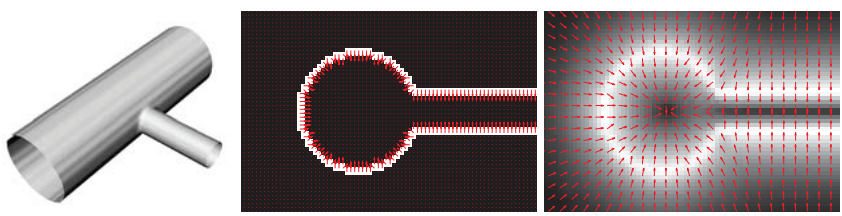
In this work, we adapt our method such that an explicit segmentation is derived and utilize it to automatically segment airway trees in CT data. The method is generally applicable and may also be used in other application domains. For evaluation of our approach, the airway trees in 20 clinical lung CT datasets were segmented and evaluation results are reported.

2 Method

Our method consists of two main steps. First, centerlines of tubular objects are extracted. Second, the tubular structures associated with these centerlines are segmented. Intermediate processing results are depicted in Figs. 2, 3, and 4; appropriate masks were used for better visualization of some intermediate results.

2.1 Gradient Vector Flow

Our method is based on the Gradient Vector Flow (GVF) [7] – an edge preserving gradient vector diffusion function – and specifically requires an appropriate initial vector field where the vectors point towards the center of the structures of interest (airways) and whose magnitude reflects an edge-likeliness (Fig. 1(b)). Therefore, the inverted local derivatives $F = -\nabla(G_\sigma \star I)$ are computed and normalized $F^n(\mathbf{x}) = \frac{F(\mathbf{x})}{|F(\mathbf{x})|} \frac{min(|F(\mathbf{x})|, F_{max})}{F_{max}}$ for every voxel $\mathbf{x} = \{x, y, z\}$, where I is the original image and G_σ is a Gaussian filter kernel at scale σ. σ and F_{max} are application specific parameters reflecting the noise level and the expected contrast.



(a) (b) (c)

Fig. 1. Illustration of the GVF using a 2D cross section of a 3D branching tubular structure. (a) Branching 3D tubular structure. (b) Initial vector field $F^n(\mathbf{x})$ (gray-value: vector magnitude; arrow: vector direction). (c) Resulting GVF field $V(\mathbf{x})$ (gray-value: vector magnitude; arrow: vector direction).

Given this initial vector field $F^n(\mathbf{x})$, the GVF is calculated which is defined as the vector field $V(\mathbf{x})$ that minimizes:

$$E(V) = \iiint_{\Omega} \mu|\nabla V(\mathbf{x})|^2 + |F^n(\mathbf{x})|^2|V(\mathbf{x}) - F^n(\mathbf{x})|^2 d\mathbf{x} \quad (1)$$

where μ is a regularization parameter. The variational formulation of the GVF keeps vectors with large magnitude nearly equal, while it produces a slowly varying field in areas with small vector magnitude. An initial vector field $F^n(\mathbf{x})$ and the GVF result $V(\mathbf{x})$ are shown in Figs. 1(b) and (c), respectively. For tubular objects, some characteristic properties can be observed. All vectors point from the boundary toward the center of the tubular objects where the vectors "collide". The vector field shows a large variation in the cross-sectional planes of the tubular objects, but almost no variation along the tubes tangent direction. In addition, the magnitudes of the vectors show a medialness property: their values decreases with increasing distance from boundaries. At the centers of the tubular objects, the magnitude almost vanishes (it not necessarily becomes zero) and forms local directional minima. Examples of the initial and the GVF vector fields for a CT dataset are shown in Fig. 2. Below we will refer to the GVF field's normalized direction as $V^n(\mathbf{x}) = V(\mathbf{x})/|V(\mathbf{x})|$ and to its magnitude as $M(\mathbf{x}) = |V(\mathbf{x})|$.

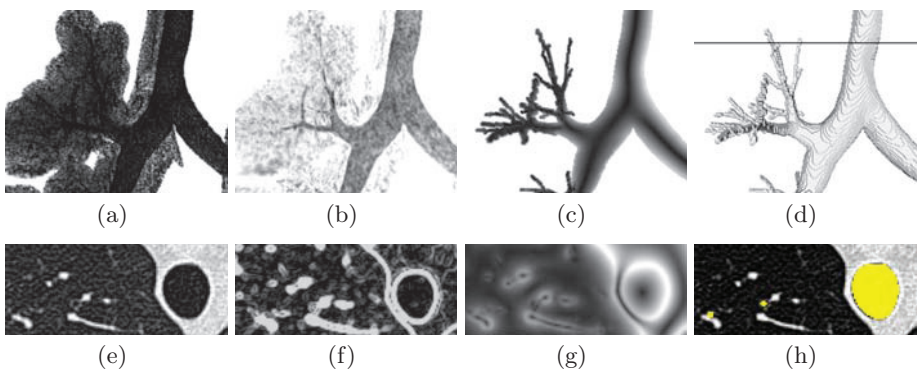


Fig. 2. Example showing some properties of the initial and the GVF's vector fields magnitudes on a real dataset. (a) Minimim Intensity Projection (MinIP) of the dataset. (b) MinIP showing the Gauss-smoothed dataset with $\sigma = 0.5$ that was used to calculate the initial gradient $F(\mathbf{x})$. (c) MinIP of the GVF magnitude $M(\mathbf{x})$ inside the segmentation result. (d) Segmentation result; the axial cutting plane used in (e)-(h) is indicated by a black line. (e) Axial slice of the dataset showing part of the trachea and some thin low contrast airways. (f) Magnitude of initial vector field $|F^n(\mathbf{x})|$ before applying the GVF. (g) Magnitude $M(\mathbf{x})$ of the GVF field. (h) Segmentation result.

In the following sections, we will show how we utilize these properties for airway centerline extraction and segmentation.

2.2 Tube Centerline Extraction

To identify tubular objects and to extract their centerlines, the Hessian matrix $H(\mathbf{x}) = \nabla V(\mathbf{x})$ with its eigenvalues $|\lambda_1| \leq |\lambda_2| \leq |\lambda_3|$ and eigenvectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ is computed for all voxels in the image to obtain the tubes cross-sectional plane spanned by $\mathbf{v}_2$ and $\mathbf{v}_3$. Based on this information, center points are identified as local directional minima in the medialness map M. In this way, center points (colored points in Fig. 3(a)). To identify those center points that are related to tubular objects, a tube-likeliness $T(\mathbf{x})$ is computed. For this purpose, a circle is fitted to the data in the tubes cross-sectional plane. The quality of the fitting determines the tube-likeliness. Compared to using a weighting of the eigenvalues of the Hessian matrix to determine a tube-likeliness [5, 6], this allows for a higher selectivity. The fitting term [12] is computed as the mean flow through the circle and depends on the radius r: $T(\mathbf{x}, r) = \frac{1}{2r\Pi} \int_{\alpha=0}^{2\Pi} \langle V(\alpha, r), D(\alpha) \rangle \, \mathrm{d}\alpha$. $V(\alpha, r)$ represents the GVFs vector at the circle point and $D(\alpha)$ defines a normal vector on the circle pointing towards its center. The integral is approximated by computing the sum over 32 discrete circle points. During the circle fitting procedure, the radius is steadily increased until the circle touches an actual edge/surface of the object. Increasing r further results in a drop of $T(\mathbf{x}, r)$ as the magnitude of the vectors drop off. Thus, the fitting is performed for increasing radii as long as the fitting term increases. The best fit determines the tube-likeliness $T(\mathbf{x})$.

Applying this procedure to the GVF field results in a tube-likeliness measure at the centerlines, as shown in Fig. 3(b). This information can be used for detection and centerline extraction of tubular objects. However, for thin low contrast airways, the response may fall off strongly, if their gradient-magnitude is too low so that they are not completely preserved in the GVF result (Figs. 2(f) and (g)). Applying the same procedure with a radius of $0.5mm$ on the initial vector field $F^n(\mathbf{x})$ allows identification of these structures as shown in Fig. 3(b), and therefore, the maximum of both responses is utilized to produce a combined tube-likeliness volume. To extract the centerlines and to discard non-tubular objects, a hysteresis thresholding with t_h and t_l is performed on the tube-likeliness volume and neighboring centerline points are linked together into centerlines. From these initial centerlines, short spurious centerlines with a length (below t_s) are discarded. In addition, centerlines with a mean tube-likeliness below t_m are removed. The resulting centerlines of the tubular objects are shown in Fig. 3(c).

2.3 Inverse Gradient Flow Tracking Tube Segmentation

After extraction of the airway centerlines, the associated image regions are segmented. In the GVF field, the vectors flow towards the centers of the airways, which correspond to the extracted centerlines (in case of tubular objects). By following the direction of the gradient vectors, each voxel can be assigned to a neighboring voxel and a path can be tracked for each voxel of the airway to its centerline. Based on this assignment and the fact that the gradient magnitude increases until the actual boundary is reached, the airways associated with the

(a) (b) (c) (d)

Fig. 3. Intermediate results of the tube centerline extraction method. The tube likeliness for the computed centerline points is shown as intensity value, and the extracted centerlines are drawn in green. (a) Tube-likeliness for larger structures. (b) Tube-likeliness for thin low contrast structures. (c) Extracted tube centerlines. (d) Segmentation result showing the size of the airways.

individual centerlines are segmented by following the gradient flow tracking path in the inverse direction. Starting from the points at the tubes center, neighboring voxels fulfilling these criteria are merged iteratively to generate a segmentation (Algorithm 1).

Algorithm 1 Inverse gradient flow tracking tube segmentation

input: GVF field $V(x)$ with direction $V^n(x)$ and magnitude $M(x)$
input: centerline points $C = \{x_1, x_2, \ldots, x_n\}$
set $S \leftarrow C$
queue $Q \leftarrow C$
while $Q \neq \{\}$ **do**
  $x \leftarrow \text{extract}(Q)$
  for each voxel $y \in Adj26(x)$ **do**
    if $y \notin S$ and $M(y) > M(x)$ and $argmin_{z \in Adj26(y)} \langle V^n(y), \overrightarrow{yz} \rangle = x$ **then**
      $S \leftarrow S \cup \{x\}$
      $Q \leftarrow Q \cup \{y\}$
    end if
  end for
end while
output: segmented tube voxels S

Examples of segmented airway branches assigned to individual centerlines generated with this algorithm are shown in Fig. 4. Note that the individually segmented parts do not "leak" into side branches of the airway tree, and the combination of all individually segmented airway branches provides a valid segmentation of the complete airway tree. Another example segmentation is shown in Fig. 2(h) where the segmentation of thin low contrast airways is clearly visible.

To discard non-airway structures and to assure a 6-connected segmentation result, the following post-processing was applied to the datasets. The extracted centerlines were dilated and added to the segmentation result to assure

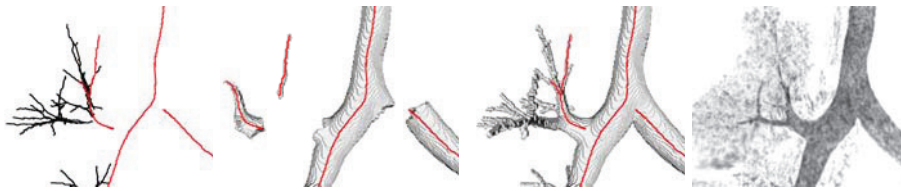


Fig. 4. Illustration of the inverse gradient flow tracking tube segmenation. (a) Extracted centerlines showing 4 selected centerlines. (b)-(e) Segmentations associated with the selected centerlines. (f) Combined results of all tubular structures. (h) MinIP of smoothed dataset for comparison.

6-connectivity. The actual airway tree was identified as the largest connected segmented component and other segmented tubular structures were discarded.

3 Evaluation

The approach was applied to 40 clinical CT datasets of the thorax (undisclosed reference segmentations) which were provided by the organizers of the "Extraction of Airways from CT 2009 (EXACT09)" workshop. The datasets were split into two sets of 20 training datasets where parameters were adapted and 20 testing datasets. The following set of parameters was used for segmentation of all 40 datasets: $\sigma = 0.5$, $F_{max} = 200$, $\mu = 5$ $t_h = 0.5$, $t_l = 0.1$, $t_s = 5$, and $t_m = 0.5$. The GVF was computed using an iterative update scheme with 500 iterations [7]. The segmentation results were sent to the organizers, who in return provided evaluation results for the 20 test cases (see Table 1). For information about how the reference segmentations were obtained and the exact definition of the used performance measures we refer to `http://image.diku.dk/exact/information.php`. On average, 63.0% of airway branches were detected with an average detected tree length of 58.4%. The mean leakage count was 5.0, and the mean false positive rate was 1.44% (median: 0.61%).

4 Discussion

The evaluation results show that our method achieves a good average airway branch count (63.0%) without any major leakage. One exception is CASE39, where some leakage occurred ($3577mm^3$).

In the following, we discuss some properties of our approach. Examples of segmentation results are shown in Fig. 5 and Fig. 6. As can be seen, the segmented surface shows a good correspondence with the image data. The tube detection/segmentation method is also capable of handling cases where the airway shape deviates from a perfectly circular or purely convex shape (e.g., trachea

Table 1. Evaluation results on the twenty test cases.

	Branch count	Branch detected (%)	Tree length (cm)	Tree length detected (%)	Leakage count	Leakage volume (mm^3)	False positive rate (%)
CASE21	114	57.3	68.5	62.0	3	35.2	0.31
CASE22	270	69.8	218.9	66.2	9	474.6	1.60
CASE23	187	65.8	134.0	51.5	5	43.5	0.20
CASE24	139	74.7	113.9	70.0	8	176.9	0.59
CASE25	158	67.5	123.1	48.8	4	98.6	0.32
CASE26	59	73.8	51.2	78.0	2	274.3	2.70
CASE27	77	76.2	58.1	71.7	4	353.5	3.06
CASE28	86	69.9	66.9	61.0	1	49.6	0.43
CASE29	120	65.2	81.2	58.8	4	118.9	0.85
CASE30	114	58.5	87.5	57.2	5	98.9	0.64
CASE31	96	44.9	70.5	40.2	1	59.8	0.34
CASE32	101	43.3	80.3	36.8	1	175.2	0.86
CASE33	117	69.6	90.4	61.5	1	32.0	0.29
CASE34	250	54.6	184.6	51.6	11	358.1	1.05
CASE35	168	48.8	110.9	35.9	5	69.8	0.30
CASE36	294	80.8	330.6	80.2	5	78.6	0.25
CASE37	112	60.5	87.9	49.5	2	102.9	0.48
CASE38	64	65.3	51.2	77.1	4	311.0	2.64
CASE39	291	56.0	250.6	61.2	13	3577.0	9.21
CASE40	225	57.8	187.7	48.5	11	959.0	2.65
Mean	152.1	63.0	122.4	58.4	5.0	372.4	1.44
Std. dev.	75.7	10.4	75.2	13.2	3.6	785.4	2.06
Min	59	43.3	51.2	35.9	1	32.0	0.20
1st quartile	96	56.0	68.5	48.8	2	59.8	0.31
Median	119	65.3	89.2	59.9	4	110.9	0.61
3rd quartile	250	73.8	187.7	71.7	9	358.1	2.65
Max	294	80.8	330.6	80.2	13	3577.0	9.21

of CASE18 in Figs. 5 and 6). Leakage into non-tubular structures is uncommon. However, in case of emphysema some leakage was observed (e.g., CASE14 in Fig. 5).

An additional advantage of our approach is that the presented centerline extraction method may be easily extended to extract the complete curve skeleton of the airway tree. This can be achieved by obtaining connections between the individual centerlines based on the GVFs medialness property as we showed in our previous work [6]. An example of the skeleton extracted with our method from the already computed GVF field is shown in Fig. 7(a). For comparison, we also extracted a skeleton with the skeletonization approach presented by Palagyi et al. [13] based on the binary segmentation obtained with our presented GVF segmentation approach. A comparison with other skeletonization approaches is provided in [6]. As can be seen, the skeleton extracted with our GVF based approach has high quality.

Regarding computation time, we utilize a GPU (graphics processing unit) implementation using the CUDA framework[1] for computation of the GVF. Other parts of the implementation are not optimized. On average, our algorithm requires 6 minutes for segmentation of the airways using an NVIDIA Tesla C1060 card processing the whole dataset. This time can be reduced to 2 minutes when the computation is restricted to an axis aligned subvolume surrounding the lung area.

5 Conclusion

In this work we presented an approach for the identification and segmentation of airway trees in CT data based on GVF. First, the method extracts centerlines of the airways. This information is then utilizes in an inverse gradient flow tracking step for the actual segmentation of individual airways. In addition, the utilized GVF field may also be used for the extraction of a high quality skeleton.

6 Acknowledgments

This work was supported in part by the Austrian Science Fund (FWF) under the doctoral program Confluence of Vision and Graphics W1209 and the Austrian Marshall Plan Foundation.

References

1. Sluimer, I., Schilham, A., Prokop, M., van Ginneken, B.: Computer analysis of computed tomography scans of the lung: A survey. IEEE Trans. Med. Imag. **25**(4) (Apr. 2006) 385–405
2. Graham, M.W., Gibbs, J.D., Higgins, W.E.: Robust system for human airway-tree segmentation. In: Proc. of SPIE Med. Imag. (Mar. 2008)

[1] `http://www.nvidia.com/object/cuda_home.html`

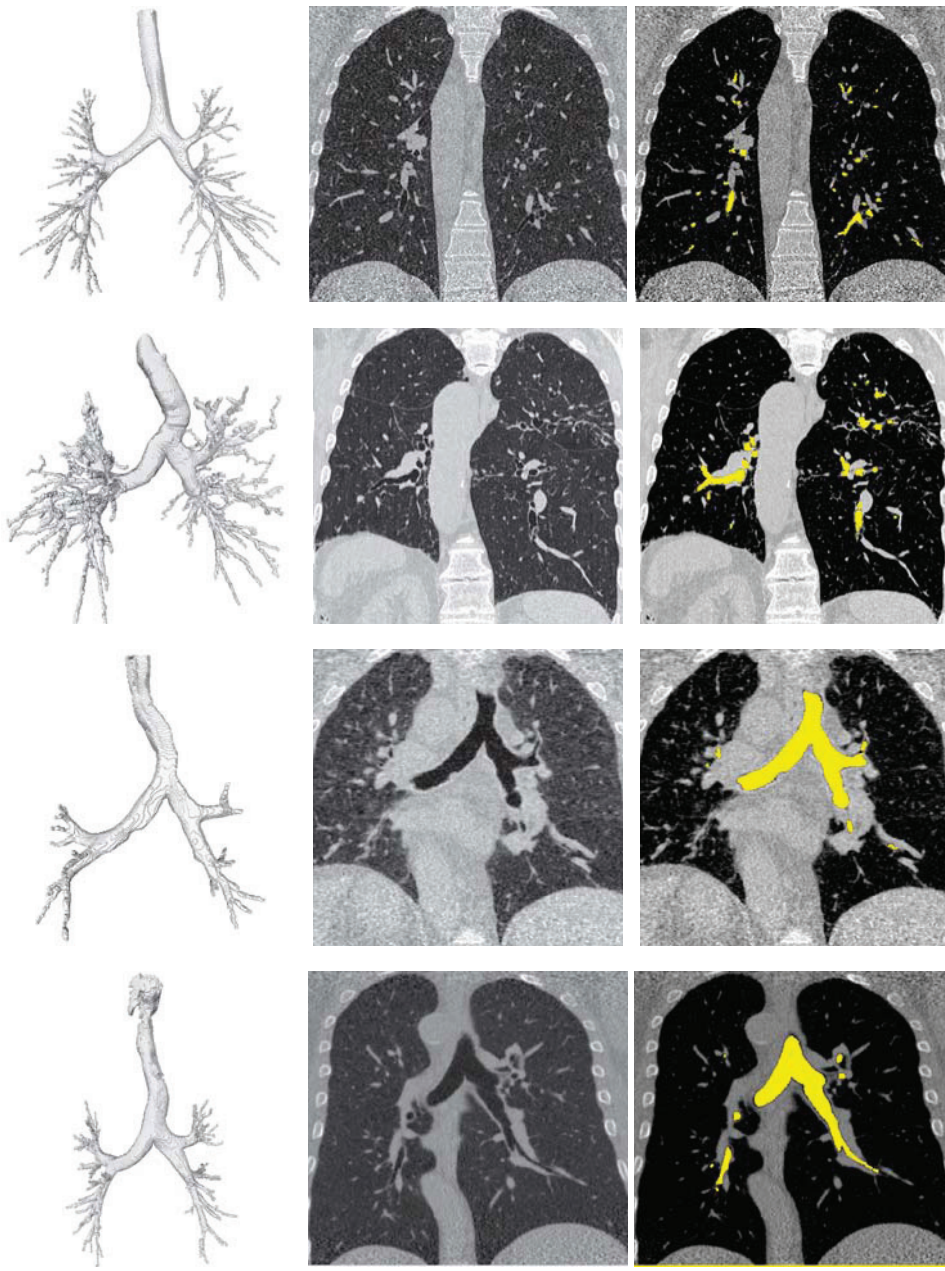


Fig. 5. Examples of segmentation results (EXACT09 database). From top to bottom: CASE03, CASE14, CASE18, CASE24. From left to right: 3D visualization of result; coronal slice showing CT data; coronal slice showing the segmentation result.

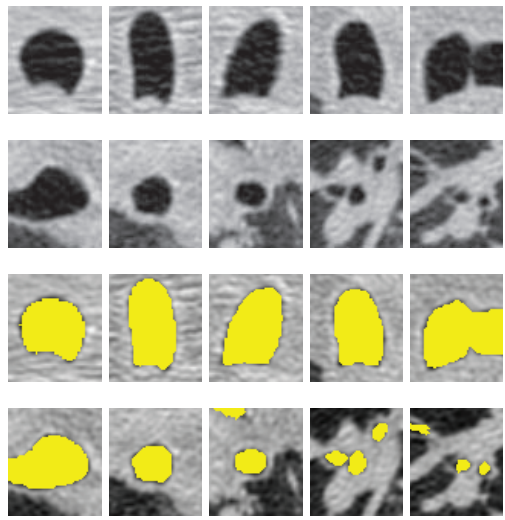
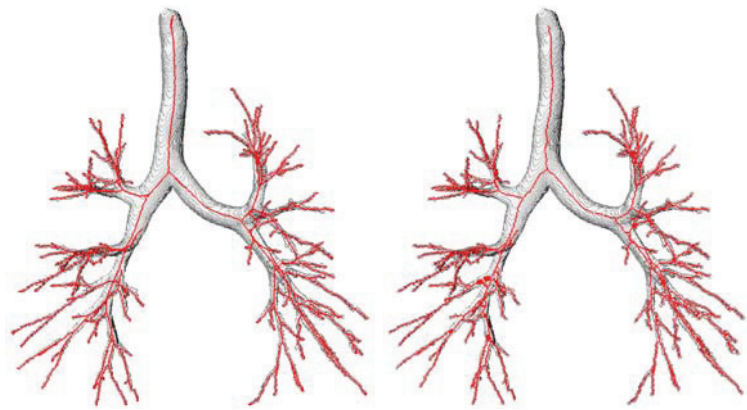


Fig. 6. Segmentation results in axial planes along a path from the root of the trachea to a distal end of an airway for a dataset where the cross-sectional profile deviates from a circular shape (CASE18; see also row 3 in Fig. 5). Note that for visualization of the CT datasets, a linear interpolation was used which resulted in a blurring of the image.



(a) (b)

Fig. 7. Curve skeletons extracted with (a) our approach presented in [6] based on the already computed GVF field and (b) the skeletonization method presented in [13] based on the binary segmentation.

3. Swift, R.D., Kiraly, A.P., Sherbondy, A.J., Austin, A.L., Hoffman, E.A., McLennan, G., Higgins, W.E.: Automatic axis generation for virtual bronchoscopic assessment of major airway obstructions. Comput. Med. Imaging Graphics **26**(2) (2002) 103 – 118

4. Tschirren, J., McLennan, G., Palágyi, K., Hoffman, E.A., Sonka, M.: Matching and anatomical labeling of human airway tree. IEEE Trans. Med. Imag. **24**(12) (Dec. 2005) 1540–1547
5. Bauer, C., Bischof, H.: A novel approach for detection of tubular objects and its application to medical image analysis. In: Proc. of DAGM. (2008) 163–172
6. Bauer, C., Bischof, H.: Extracting curve skeletons from gray value images for virtual endoscopy. In: Proc. of MIAR. (2008) 393–402
7. Xu, C., Prince, J.L.: Snakes, shapes, and gradient vector flow. IEEE Trans. on Image Processing **7**(3) (Mar. 1998) 359–369
8. Yu, Z., Bajaj, C., Bajaj, R.: Normalized gradient vector diffusion and image segmentation. In: Proc. of ECCV. (2002) 17–530
9. Chuang, C.H., Lie, W.N.: A downstream algorithm based on extended gradient vector flow field for object segmentation. IEEE Trans. Med. Imag. **13**(10) (Oct. 2004) 1379–1392
10. Schmugge, S.J., Keller, S., Nguyen, N., Souvenir, R., Huynh, T., Clemens, M., Shin, M.C.: Segmentation of vessels cluttered with cells using a physics based model. In: Proc. of MICCAI. (2008) 127–134
11. Li, G., Liu, T., Tarokh, A., Nie, J., Guo, L., Mara, A., Holley, S., Wong, S.: 3D cell nuclei segmentation based on gradient flow tracking. BMC Cell Biology **8**(40) (2007)
12. Krissian, K., Malandain, G., Ayache, N., Vaillant, R., Trousset, Y.: Model-based detection of tubular structures in 3D images. CVIU **2**(80) (2000) 130–171
13. Palágyi, K., Sorantin, E., Balogh, E., Kuba, A., Halmai, C., Erdohelyi, B., Hausegger, K.: A sequential 3D thinning algorithm and its medical applications. In: IPMI, Springer-Verlag Heidelberg (2001) 409–415

Airway Tree Reconstruction Based on Tube Detection

Christian Bauer[1,2], Thomas Pock[1], Horst Bischof[1], and Reinhard Beichel[2,3,4]

[1] Inst. for Computer Graphics and Vision, Graz University of Technology, Austria
[2] Dept. of Electrical and Computer Engineering
[3] Dept. of Internal Medicine
[4] The Iowa Institute for Biomedical Imaging
The University of Iowa, Iowa City, IA 52242, USA
{cbauer, pock, bischof}@icg.tu-graz.ac.at
reinhard-beichel@uiowa.edu

Abstract. We present an automated approach for airway tree reconstruction from CT images. Our approach performs an initial identification of tubular structures, followed by a reconstruction of the airway tree. During the reconstruction step, tubular objects that are part of the airway tree are identified and linked together based on prior knowledge about the structure of human airway trees. A major advantage of our approach is that it handles local disturbances robustly, as demonstrated by our experiments.

1 Introduction

Segmentation of airway trees in lung CT data is a prerequisite for several clinical applications including diagnosis and monitoring of lung disease or surgical planning. To facilitate such applications, it is highly desirable to have an automated airway segmentation method which is robust. The main challenges in the context of airway tree segmentation are: noise, inhomogeneous appearance of the airway wall due to partial volume effects, motion artifacts, or lung disease (e.g., emphysema).

In the literature, several airway tree segmentation methods have been presented. An overview is given in the survey of Sluimer et al. [1], and newer approaches are discussed in [2]. Frequently, region growing or front propagation methods [3–5] are utilized that make some assumptions about the density (gray-values) of airways in CT data. More sophisticated variants of front propagation methods try to avoid segmentation errors by constantly analyzing local segmentation results and by adapting parameters accordingly. Different approaches have been proposed [3, 6, 4] that utilize local information like radius or branching angle. To enhance robustness, compared to methods that solely rely on density/gray-value information for airway segmentation, some approaches focus on airway candidate detection using mathematical morphology [7], template matching techniques [8], or voxel classification based on different image descriptors [9, 10].

Many of the available approaches have problems in dealing with local disturbances (e.g., motion artifacts) or pathology (e.g., obstructed airway) which frequently results in incomplete airway segmentations. Graham et al. [2] addressed this problem by building an airway tree from candidate airway branch segments by computing connection costs between branches and using graph theoretic approaches to extract the airway tree [2]. However, their method does not consider possible connections in the context of the complete airway tree structure which can lead to false connections.

In this work, we present an automated approach for the reconstruction of airway trees that is robust against local disturbances which can result from disease or imaging artifacts, for example.

2 Method

Our method consists of two main processing steps. First, all tubular structures are identified in the dataset. Second, the airway tree is reconstructed from these tubular objects by utilizing knowledge about the tree structure. During this step, tubular objects that are not related to airways are discarded. In addition, tubular objects that are part of the airway tree are linked together. Consequently, potential gaps between airway branches, that may occur due to disturbances (e.g., imaging artifacts), are closed. Fig. 1 illustrates the individual processing steps by showing intermediate results.

2.1 Detection and Description of Tubular Objects

Airways form tubular tree structures that can be detected by utilizing a tube detection filter (TDF), because in CT data, the airway branches appear as elongated structures with varying diameter that are darker than the surrounding tissue. To detect and describe these structures, a TDF combined with a centerline extraction method based on a ridge traversal is utilized.

The utilized multi-scale TDF extends the works presented in [11, 12]. Specifically, our method achieves an increased robustness and minimizes artifacts by utilizing additional constraints. Therefore a radius dependent tube-likeliness measure $R(\mathbf{x}, r)$ is computed for every point $\mathbf{x}$ in the image domain for a set of different radii. To obtain $R(\mathbf{x}, r)$ for a given radius r, an offset medialness function and an adaptive threshold are used that utilize two different scales simultaneously: $\sigma_{\mathcal{H}} = r$ and $\sigma_{\mathbf{B}} = r^{\eta}$ with $0.0 \leq \eta \leq 1.0$. While the scale $\sigma_{\mathcal{H}}$ is large and and is used to cover the whole structure, $\sigma_{\mathbf{B}}$ is smaller and is utilized for obtaining boundary information. First, the Hessian matrix $\mathcal{H}(\mathbf{x}) = \sigma_{\mathcal{H}}^2 \left[(\partial^2 I^{(\sigma_{\mathcal{H}})})/(\partial_{x_i}\partial_{x_j})\right]$ with its associated eigenvalues $|e_1| \geq |e_2| \geq |e_3|$ and eigenvectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ is computed, where $I^{(\sigma)} = G_{\sigma} \star I(x)$ corresponds to the original image at scale σ. Based on this information, "dark" structures (low density) are identified ($e_1 > 0$ and $e_2 > 0$). For these points, an offset medialness function is computed based on boundariness contributions b_i which are sampled along a circle in the tubes cross sectional plane defined by $\mathbf{v}_1$ and

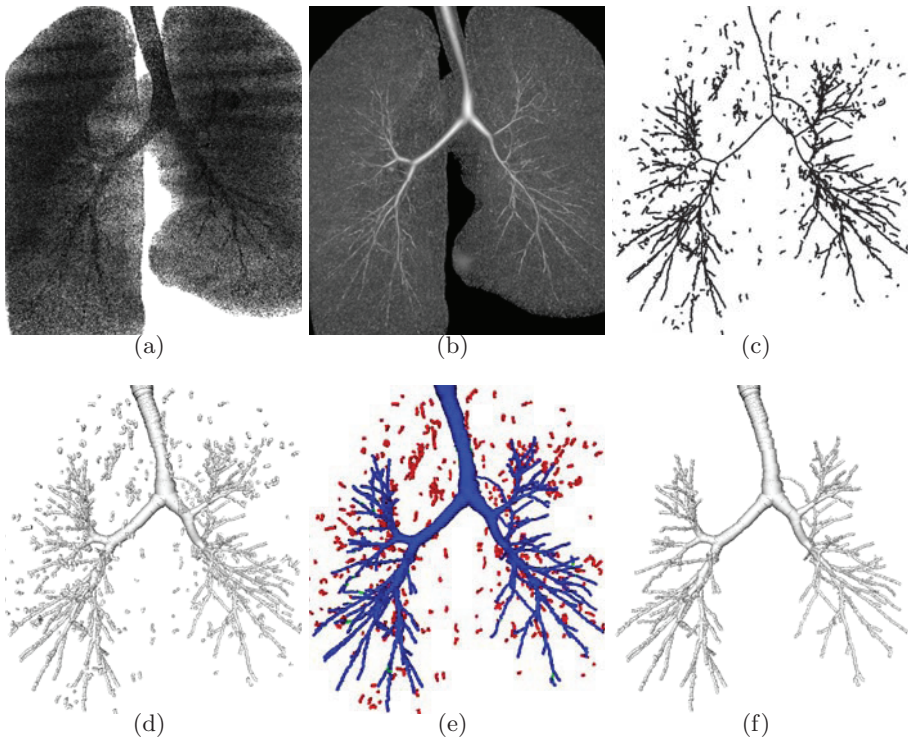


Fig. 1. Illustration of the processing steps of our airway tree reconstruction approach. (a) Volume rendering of the utilized dataset. (b) Tube detection filter response. (c) Centerlines of initially extracted tubular objects. (d) Initially extracted tubular objects with associated radii/tangent directions. (e) Tree reconstruction step showing the identified tubular objects belonging the airway tree (blue), the discarded tubular objects (red), and the closed gaps (green). (f) Reconstructed airway tree.

$\mathbf{v}_2$. The boundariness contributions are calculated at $N = \lfloor 2\pi r + 1 \rfloor$ points with varying angle $\alpha_i = (2\pi i)/N$ (N is limited to the range 16-100). The individual boundariness values $\mathbf{B}(\mathbf{x}) = \sigma_{\mathbf{B}} \nabla I^{(\sigma_{\mathbf{B}})}(\mathbf{x})$ represent a measure for the contribution of the gradient in the radial direction $\mathbf{v}_{\alpha_i} = \cos(\alpha_i)\mathbf{v}_1 + \sin(\alpha_i)\mathbf{v}_2$ of the tube: $b_i = |B(\mathbf{x} + r\mathbf{v}_{\alpha_i})\mathbf{v}_{\alpha_i}|$. Based on the average $\bar{b}(\mathbf{x}, r) = \frac{1}{N}\sum_{i=0}^{N-1} b_i$ and the variance $s^2(\mathbf{x}, r) = \frac{1}{N}\sum_{i=0}^{N-1}\left(b_i - \bar{b}\right)^2$ of these boundariness samples, the offset medialness function is computed: $R_0(\mathbf{x}, r) = \bar{b}(\mathbf{x}, r)(1 - s^2(\mathbf{x}, r)/\bar{b}(\mathbf{x}, r)^2)$. The second term that includes the variance allows suppressing responses for not circular symmetric structures. To avoid wrong responses that could occur near edges, an adaptive thresholding scheme based on the gradient magnitude is used to suppress such responses, and the final medialness response for a given radius is computed: $R(\mathbf{x}, r) = \max\{R_0(\mathbf{x}, r) - \sigma_{\mathcal{H}}|\nabla I^{(\sigma_{\mathcal{H}})}(\mathbf{x})|, 0\}$. The mutli-scale response $R_{multi}(\mathbf{x}) = \max_{r_{min} \le r \le r_{max}} \{R(\mathbf{x}, r)\}$ is obtained as the maximum

response over all radii, which also yields the associated radius r and tangent direction $t = \mathbf{v}_3$ for each point.

An example of the TDF response is shown in Fig. 1(b). The response increases towards the tube center. Consequently, the tubular objects form ridges. As can be seen in Fig. 1(b), the filter enhances dark tubular structures (airways) without producing artifacts or responses to other image structures like the lung surface, for example. However, the response decreases in proximity of junctions, and is lower for thin low contrast airways. Consequently, separating the tubular structures from noise based on a simple global thresholding would not be robust. To address this issue, a ridge traversal procedure with a hysteresis thresholding scheme is utilized to extract centerlines. The centerlines are then analyzed and noise induced responses are discarded. The procedure requires a starting point $\mathbf{x}_0$ for every ridge and an estimate of its tangent direction which is provided by the TDF. All local maxima in the TDF response with a value $R_{multi}(\mathbf{x}) > t_{high}$ are considered as starting points and used for initialization of the traversal. Starting from a given point $\mathbf{x}_0$, the ridge is traversed in both directions $\mathbf{t}_0$ and $-\mathbf{t}_0$. Given a point on the ridge $\mathbf{x}_i$, the next point $\mathbf{x}_{i+1}$ on the ridge in the traversal direction $\mathbf{t}_i$ is chosen as the local neighbor $\mathbf{x}_i^n$ with the highest value of $R_{multi}(\mathbf{x}^n)$ that satisfies $\overrightarrow{\mathbf{x}_i\mathbf{x}_i^n} \cdot \mathbf{t}_i > 0$. The traversal direction $\mathbf{t}(\mathbf{x}_{i+1}) = \text{sign}(\overrightarrow{\mathbf{x}_i\mathbf{x}_{i+1}} \cdot \mathbf{t}(\mathbf{x}_{i+1}))\mathbf{t}(\mathbf{x}_{i+1})$ is updated to maintain the correct direction, and the procedure is repeated as long as $R_{multi}(\mathbf{x}_{i+1})$ stays above a given threshold t_{low} or an already traversed point is found. Thus, object centerlines $\{l^j\}_{j=1}^m$ are extracted for all tubular objects, consisting of an ordered set of points $\{l^j\}_{j=1}^m$ with associated radius r_i^j and tangent directions $\mathbf{t}_i^j$. To discard short spurious responses (noise), all centerlines with an accumulated TDF response below t_{conf} are discarded. For the remaining centerlines the radius and tangent directions are re-estimated by averaging over the ± 5 local neighbors along the centerline. The centerlines are split into subparts at local angles larger than 90° to guarantee that at furcations at least one of the centerlines has an endpoint. This is a necessary prerequisite for the next processing step. The angle is determined based on the ± 5 local neighbors along the centerline.

Figs. 1(c) and (d) depict the resulting descriptions of the identified tubular objects. Fig. 1(c) shows only the centerline information, while in Fig. 1(d) also the associated radius and tangent direction are displayed using a cylinder with appropriate orientation and radius. As can be seen, major parts of the airway tree can be extracted with this approach. However, two problems remain that have to be addressed. First, the centerlines of the tubular objects may break up at junctions or in disturbed regions (e.g., motion artifacts). Second, some false positive responses caused by other low density (dark) tube-like structures that are not airways are also present.

2.2 Tree Reconstruction

For reconstruction of the airway tree, tubular objects that are part of the actual airway tree need to be identified, and all other unrelated tubular objects must

be discarded. For this purpose, the structure and the relation between the identified tubular objects is analyzed. In addition, prior knowledge is utilized for the reconstruction. Starting from the trachea, the airway tree furcates recursively. At furcations, the radius of the child branches decreases. In addition, no abrupt changes of the tangent direction occur. Our method incorporates this knowledge. Beginning with the trachea, tubular objects that follow this pattern are merged.

During the tree reconstruction process, a graph-based representation of the whole tree is derived, describing the tree structure. In this graph, nodes represent branchpoints and edges correspond to tubular objects connecting these branch points. For all tubular objects l^j, the average radius r^j is determined and the proximal/distal direction of the tube element $d^j \in \{+1, -1\}$; $d^j = +1$ if the direction is from the first centerline point to the last or $d^j = -1$ otherwise. In addition, we define the angles $\alpha^l = \angle\big(\overrightarrow{\mathbf{x}_k^l \mathbf{x}_i^j}, d^l t_k^l\big)$ and $\alpha^j = \angle\big(\overrightarrow{\mathbf{x}_k^l \mathbf{x}_i^j}, d^j t_i^j\big)$ and the distance $d = \max\big(0, |\overrightarrow{\mathbf{x}_k^l \mathbf{x}_i^j}| - r_k^l\big)$ between points $\mathbf{x}_i^j$ and $\mathbf{x}_k^l$ of the tubular objects which are required for the tree reconstruction process. (Fig. 2).

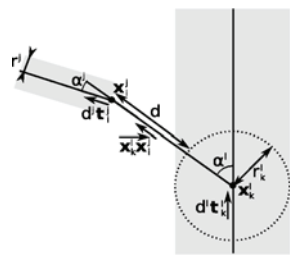


Fig. 2. Branch angles and the distance between tubular structures as used by the tree reconstruction.

Starting from the trachea, the airway tree is reconstructed by iteratively merging unconnected tubular objects. Therefore, connection confidences are calculated between endpoints of unconnected tubular objects and the current airway tree: $c(\mathbf{x}_i^j, \mathbf{x}_k^l) = exp(-\alpha^j/2\rho^2)(1 + d/r^j)$. This confidence decreases with increasing distance and increasing angle. In addition, to be considered a valid connection, the following constraints have to be fulfilled:

1. the branching angle must not be too large ($\alpha^j \leq \gamma_a$ and $\alpha^l \leq \gamma_a$),
2. the radius must not increase ($r^j \leq \gamma_r r_m$; where r_m is the smallest radius on the whole path from the trachea),
3. the connection distance must not be too large ($d \leq \gamma_d$), and
4. a minimum connection confidence is required ($c \geq \gamma_c$).

After determination of the connection with the highest confidence that fulfills all above described constraints, the associated tube element is added to the airway tree and the structural representation is updated. In addition, the identified connection is also added to the airway tree by using a linearly interpolated

path. By applying the method recursively, gaps between unconnected airway branches are closed and a connected skeleton of the whole airway tree is obtained, which includes additional radius and tangent direction for each centerline point (Figs. 1(e) and (f)).

2.3 Preprocessing Steps for Automated Airway Tree Reconstruction

The above outlined principle is general applicable. To fully automate the approach for airway detection, the input CT dataset is preprocessed to discard non lung tissue and to restrict the search area for tubular objects. Therefore a rough lung mask is generated by using thresholding ($< -700HU$), connected component analysis, and morphological closing with a ball structuring element with a radius of 10 voxels. All voxels outside this lung mask or with a value larger than $-700HU$ in the original dataset are set to $-700HU$. The resulting dataset was used as input for our method. The trachea is identified automatically by searching for the largest tubular object located at the center of the volume.

3 Evaluation

Our approach was applied to 40 clinical datasets (with undisclosed gold standard) which were provided by the organizers of the "Extraction of Airways from CT 2009 (EXACT09)" workshop (`http://image.diku.dk/exact`). The datasets were split in two groups of 20 training datasets, where the parameters have been adapted and 20 testing datasets. For information about how the reference segmentations were obtained and the exact definition of the used performance measures we refer to `http://image.diku.dk/exact/information.php`.

The following parameter were used to process the test datasets. The tube detection was performed on 15 radius steps on a logarithmic scale between radii $0.25mm$ and $10mm$ with $\eta = 0.7$ (the variance term of the boundariness samples in the offset medialness function was omitted for radii $\leq 0.5mm$); $t_{high} = 35$, $t_{low} = 25$ and $t_{conf} = 150$ for the centerline extraction; $\rho = 0.5$, $\gamma_a = 90°$, $\gamma_r = 1.3$, $\gamma_d = 40mm$ and $\gamma_c = 0.1$ for the tree reconstruction.

For evaluation, binary volume datasets were required that contain a single 6-connected airway structure. Our airway tree reconstruction method produces a 26-connected airway tree skeleton with corresponding radius information. Thus, to obtain a binary volume dataset, we performed an inverse distance transformation to obtain a rough segmentation and dilated the so obtained reconstruction to assure 6-connectivity. The generated segmentations were sent to the organizers, who in return provided evaluation results. Table 1 summarizes the evaluation results for the 20 testing datasets. On average, 57.9% of airway branches were detected with an average detected tree length of 55.2%. The mean leakage count was 6.5, and the mean false positive rate was 2.44% (median: 1.41%).

Table 1. Evaluation results on the twenty test cases.

	Branch count	Branch detected (%)	Tree length (cm)	Tree length detected (%)	Leakage count	Leakage volume (mm^3)	False positive rate (%)
CASE21	100	50.3	54.6	49.4	1	102.5	1.09
CASE22	296	76.5	258.7	78.3	23	1311.2	3.91
CASE23	232	81.7	201.0	77.2	14	700.5	2.81
CASE24	148	79.6	121.0	74.4	5	288.7	1.18
CASE25	157	67.1	134.0	53.2	9	1693.0	6.08
CASE26	39	48.8	29.4	44.7	4	254.1	5.13
CASE27	38	37.6	27.6	34.1	1	56.8	1.08
CASE28	70	56.9	50.9	46.5	2	8.6	0.09
CASE29	118	64.1	83.6	60.5	3	222.6	1.56
CASE30	91	46.7	70.2	46.0	1	47.6	0.32
CASE31	100	46.7	73.8	42.0	1	64.7	0.38
CASE32	125	53.6	118.6	54.4	2	34.8	0.14
CASE33	117	69.6	99.1	67.4	15	931.2	7.56
CASE34	296	64.6	233.5	65.3	6	167.3	0.49
CASE35	187	54.4	133.2	43.1	2	123.3	0.57
CASE36	239	65.7	283.0	68.7	8	661.1	2.12
CASE37	96	51.9	82.8	46.6	5	360.2	2.25
CASE38	40	40.8	30.8	46.3	2	79.6	1.25
CASE39	210	40.4	198.1	48.4	8	745.7	2.61
CASE40	237	60.9	219.6	56.7	17	3678.0	8.07
Mean	146.8	57.9	125.2	55.2	6.5	576.6	2.44
Std. dev.	81.8	13.0	80.3	12.8	6.3	864.9	2.46
Min	38	37.6	27.6	34.1	1	8.6	0.09
1st quartile	91	46.7	54.6	46.0	2	64.7	0.49
Median	122	55.6	108.8	51.3	5	238.3	1.41
3rd quartile	237	69.6	219.6	68.7	14	931.2	5.13
Max	296	81.7	283.0	78.3	23	3678.0	8.07

4 Discussion

In this paragraph, we discuss properties of our approach based on two examples. In Fig. 3, a case with emphysema is shown. Several structures that are similar to airways are present in the image data. They are hard to distinguish solely based on gray-value appearance. Without utilizing prior knowledge about the structure of airway trees, a correct segmentation is hard to achieve, and simple algorithms will likely show leakage. As shown in Fig. 3(b), our approach initially also identifies some of these structures as tubular objects. However, the constraints incorporated into the tree reconstruction step successfully prevent that such structures are added to the airway tree (Fig. 3(d)). In the example shown in Fig. 4, a tumor infiltrates the airway wall and blocks one of the lower airway branches completely, such that the airway tree appears to be unconnected in the image data. Contrary to region growing or front propagation approaches, our approach identifies the unconnected airways and allows us to link them together. This ability enables our method to handle local disturbances robustly.

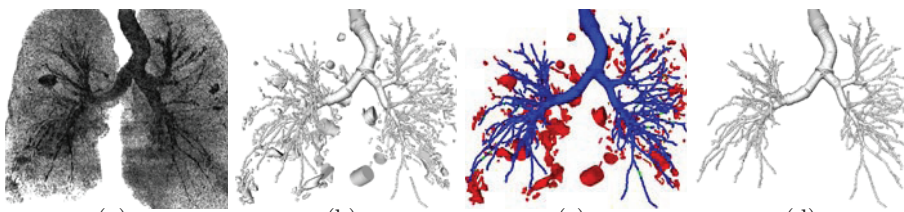


Fig. 3. Emphysema lung CT data. (a) Volume rendering of the dataset. (b) Identified tubular objects. Note that some of the "pathological" structures have been detected as tubular objects. (c) Tree reconstruction step showing the identified tubular objects belonging to the airway tree (blue), the discarded tubular objects (red), and the closed gaps (green). (d) Reconstructed airway tree.

Compared to other methods, we achieve a good trade-off between "branch count"/"tree length" and leakage. Fig. 5 depicts the three cases with the largest "leakage volume". Two points can be observed: i) the majority of the "leaks" detected by the workshop organizers are due to surface representation inaccuracies (Figs. 5(c) and (d)) and ii) blobs are included in some airway segmentations (Figs. 5(a) and (b)). Point i) can be explained as follows. Our approach produces a description of the airway tree on a structural level (centerline points, radius, tangent direction), but not a voxel or sub-voxel accurate segmentation of the inner and/or outer airway wall(s). In addition, since results were required to be 6-connected, we decided to dilate our results, which clearly negatively impacts leakage performance metrics. As shown in Fig. 5(d), voxels are classified as leakage, if the segmentation is thicker than the reference segmentation even if its structure is correct. Point ii) was caused by a bug in the software that

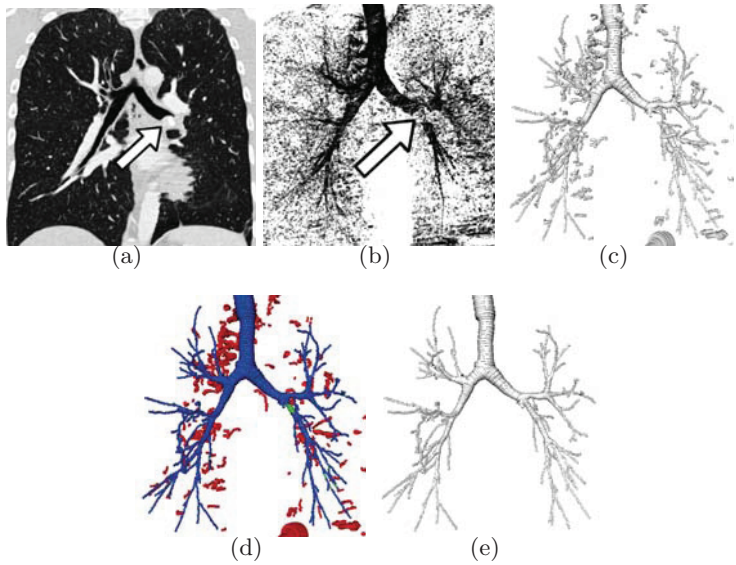


Fig. 4. Dataset with a large tumor that infiltrates and blocks one airway branch completely (indicated by the arrow). Note the gap between the airway branches at the tumor region. (a) Coronal slice of the dataset. (b) Volume rendering of the dataset. (c) Identified tubular objects showing also a gap between airway branches. (d) Tree reconstruction step showing the identified tubular objects belonging the airway tree (blue), the discarded tubular objects (red), and the closed gaps (green). (e) Reconstructed airway tree.

transformed our structural airway description into the binary volume dataset which was discovered after the evaluation. In the future we plan to improve our method by adding a surface segmentation step. The generated structural information will be utilized as a shape prior to constrain a consecutive segmentation step. The presented methods for tube detection, centerline extraction, and tree reconstruction are generic and can be utilized in other application domains.

Compared to simple airway segmentation approaches (e.g. region growing) our approach is more complex. However, TDFs as used with our approach are highly parallelizeable and well suited for a GPU (graphics processing unit) based implementation. Using a CUDA[1] based implementation of the TDF running on an NVIDIA Tesla C1060 card, the TDF response is computed on average in approximately 30 seconds per scale for the whole dataset. Using axis aligned subvolumes that just contain the lung area, computation time can be reduced to about 10 seconds per scale. The Processing time for preprocessing, ridge traversal, and tree reconstruction combined are about 30 seconds. Overall, computation time using such a subvolume containing only the lung is about 3 minutes.

[1] `http://www.nvidia.com/object/cuda_home.html`

(a) (b) (c) (d)

Fig. 5. Examples of reconstructed airways with a high "leakage volume". (a) CASE40: "leakage count"=17; "leakage volume"=3678.0 mm^3. (b) CASE25: "leakage count"=9; "leakage volume"=1693.0 mm^3. (c) CASE22: "leakage count"=23; "leakage volume"=1311.2 mm^3. (d) CASE22 comparison to gold standard: voxels classified as leakage (red), correct voxels (green), missed airways (yellow).

5 Conclusion

In this work we presented an automated approach for the reconstruction of airway trees from CT datasets. The approach utilizes local appearance information in combination with prior knowledge about the structure of airway trees. It first identifies tubular objects which are then grouped together to form an airway tree. As demonstrated on examples, our approach allows to robustly deal with cases where parts of the airway tree are locally disturbed.

6 Acknowledgments

This work was supported in part by the Austrian Science Fund (FWF) under the doctoral program Confluence of Vision and Graphics W1209 and the Austrian Marshall Plan Foundation.

References

1. Sluimer, I., Schilham, A., Prokop, M., van Ginneken, B.: Computer analysis of computed tomography scans of the lung: A survey. IEEE Trans. Med. Imag. **25**(4) (Apr. 2006) 385–405
2. Graham, M.W., Gibbs, J.D., Higgins, W.E.: Robust system for human airway-tree segmentation. In: Proc. of SPIE Med. Imag. (Mar. 2008)
3. Schlathalter, T., Lorenz, C., Carlsen, I., Reinisch, S., Deschamps, T.: Simultaneous segmentation and tree reconstruction of the airways for virtual bronchoscopy. In: Proc. of SPIE Med. Imag. (2002) 103–113
4. van Ginneken, B., Baggerman, W., van Rikxoort, E.M.: Robust segmentation and anatomical labeling of the airway tree from thoracic CT scans. In: Proc. of MICCAI, New York, USA (Sep. 2008) 219–226
5. Fabianska, A.: Two-pass region growing algorithm for segmenting airway trees from MDCT chest scans. Comput. Med. Imaging Graphics **23**(11) (Nov. 2009) 1353–1364

6. Tschirren, J., McLennan, G., Hoffman, E., Sonka, M.: Intrahotacic airway trees: Segmentation and airway morhoplogy analysis from low-dose CT scans. IEEE Trans. Med. Imag. **25**(4) (Apr. 2006) 385–405
7. Fetita, C.I., Preteux, F., Beigelman-Aubry, C., Grenier, P.: Pulmonary airways: 3-D reconstruction from multiscale CT and clinical investigation. IEEE Trans. Med. Imag. **23**(11) (Nov. 2004) 1353–1364
8. Swift, R.D., Kiraly, A.P., Sherbondy, A.J., Austin, A.L., Hoffman, E.A., McLennan, G., Higgins, W.E.: Automatic axis generation for virtual bronchoscopic assessment of major airway obstructions. Comput. Med. Imaging Graphics **26**(2) (2002) 103 – 118
9. Sato, Y., Nakajima, S., Atsumi, H., Koller, T., Gerig, G., Yoshida, S., Kikinis, R.: 3D multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. In: CVRMed-MRCAS. (1997) 213–222
10. Lo, P., de Brujine, M.: Voxel classification based airway tree segmentation. In: Proc. of SPIE Med. Imag. (Mar. 2008)
11. Krissian, K., Malandain, G., Ayache, N., Vaillant, R., Trousset, Y.: Model-based detection of tubular structures in 3D images. CVIU **2**(80) (2000) 130–171
12. Xu, M., Pycock, D.: A scale-space medialness transform based on boundary concordance voting. J. Math. Imaging Vision **11**(13) (Dec 1999) 277–299

A morphological-aggregative approach for 3D segmentation of pulmonary airways from generic MSCT acquisitions

Catalin Fetita[1], Margarete Ortner[1], Pierre-Yves Brillet[2], Françoise Prêteux[1], and Philippe Grenier[3]

1 Institut TELECOM / Telecom SudParis, Evry, France,
2 Université Paris 13, Avicenne Hospital, Bobigny, France,
3 Université Paris 6, Pitié-Salpêtrière Hospital, Paris, France

Abstract. Three-dimensional segmentation of airways from multi-slice computed tomography (MSCT) is a key point in the development of computer-aided tools for respiratory investigation. The expected benefits are related to diagnosis improvement of airway pathologies, preoperative planning and follow-up. The segmentation issue becomes even more challenging with regard to the high variability of the MSCT image acquisition in clinical practice due to the different CT scanners used and the various protocols (mainly at low dose). This paper develops a generic and automated 3D airway segmentation approach able to deal with a large spectrum of MSCT protocols by exploiting a combined morphological-aggregative methodology. The proposed method was independently assessed by an external group of medical experts in the context of a segmentation challenge, on a database consisting of 20 thorax MSCT datasets. This database included acquisitions from several clinical centers equipped with different CT scanners and using various protocols. The evaluation results show a good performance of the developed approach in terms of airway segments detection accuracy, in the context of highly variable MSCT input data.

1 Introduction

Many approaches dedicated to the segmentation of airways from CT data have been developed in the last decade [1–9], but today there is no independent evaluation of their robustness and accuracy with respect to the variability of the clinical protocols and the CT devices used in clinical practice.

This paper develops an automated and generic method for 3D airway reconstruction from MSCT and assesses it by an independent expert consortium on a shared database (http://image.diku.dk/exact/) specially built-up to match most of the current situations met in clinical routine. The segmentation scheme and the options used for the adaptation to the MSCT protocols are presented in detail in Section 2. The independent evaluation results and an overall discussion are given in Section 3.

2 Method

The approach presented in this paper is an extension to low-dose MSCT acquisitions of the technique developed in [9]. It consists of two main steps, airway candidates selection and distal airway reconstruction, which are discussed in the following. The method is applied to grayscale windowed MSCT data obtained from the original Hounsfield Unit (HU) values using the LUNG window [-1000, 250] HU. During data conversion, the scan orientation is automatically corrected by comparing the surface of the body cross-section on the first/last ordered images in the DICOM sequence. The lungs are detected on a coronal projection (volume rendering) of the DICOM volume by means of morphological operators and the first/last images of the sequence are set-up according to the lungs extent.

2.1 Airway candidates selection

The selection of the airway candidates exploits the *flood size-drain leveling (FSD)*[1] morphological operator which is defined as follows. Let $f : X \subset \Re^3 \rightarrow \Re$ denote the 3D image relief, and $g_n : X \rightarrow \Re$ a constraint function derived from f by means of an increasing operator:

$$g_n = \min(f \oplus B_n,\ f \oplus \check{B}_n), \tag{1}$$

where B_n denotes an upstream neighborhood of size n, $\check{B}_n$ the symmetric set of B_n with respect to its origin, and $\oplus$ the Minkowski addition. Note that $g(x) \geq f(x), \forall x \in X$. The flood size-drain leveling of f "floods" the f relief up to a "draining" level imposed by the size of the f "basins" (Fig. 1), and can be expressed by means of the *grayscale reconstruction by erosion* [10] operator as:

$$FSD_f^n = \mathcal{R}^\epsilon{}_f(g_n) = \ominus_f^\infty(g_n), \tag{2}$$

where $\ominus_f^\infty(g)$ denotes the geodesic erosion of g within f, of infinite size, and g_n is given by eq. 1.

Let us consider the image f in Fig. 1(a), top row, and its interpretation in terms of topographic relief, bottom row. f is a two-dimensional simulation of a pulmonary field, where the central structure represents the airway tree, and the small-size valleys denote other low-regions in the lung parenchyma. The idea exploited in [9] for airway candidate selection consists of applying the FSD_f^n operator in a multiresolution scheme, first using a small n-value to "flood" small-size "valleys" disconnected from valleys of size larger than n (Fig. 1(b)), then using a n-value equal to the size of the trachea, in order to "flood" the airway region (Fig. 1(c)). Airway candidates result from an adaptive thresholding of the difference between the filtered data (Fig. 1(d)).

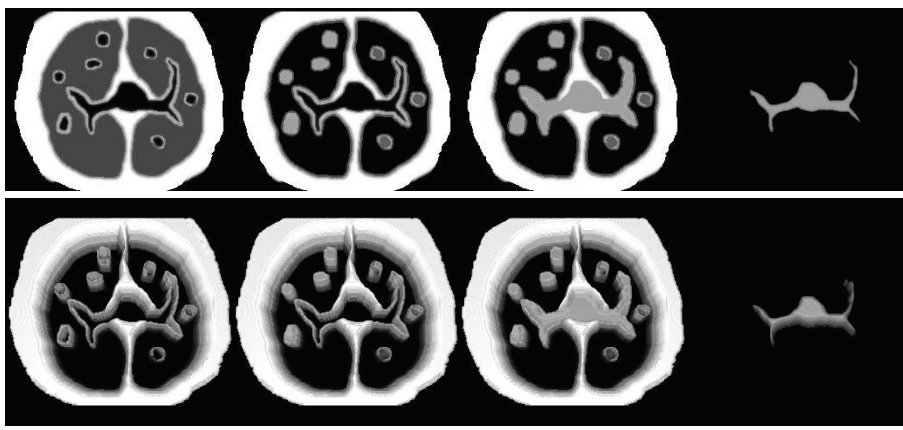
Note that, the FSD ability to "flood" a given region in the image depends on two factors: the region topography should be basin-like and its size has to be smaller than the FSD size. In low-dose MSCT data, noise and partial volume effect may induce local intensity reduction at the level of the airway wall,

[1] previously introduced as *sup-constrained connection cost* [9]

which in terms of topographic image interpretation is equivalent with a wall interruption (Fig. 2(a)). In this case, the FSD operator will fail to select the region surrounded by the airway wall, as this zone is no more a "basin" in the image relief (Fig. 2(b)). In order to preserve the selection property required, we define an additional constraint related to the size m of the largest accepted interruption in the basin's wall. An extended operator results, the *closing FSD*, given by:

$$C_m FSD_f^n = \mathcal{R}^{\epsilon}{}_{f \bullet S_m}(g_n), \tag{3}$$

where $f \bullet S_m$ denotes the morphological closing of f by a spherical structuring element of radius m. Note that, a generalization of such operator can be obtained by replacing the morphological closing with any function $h \geq f$. By applying the CSFD operator with a small m-size (usually $m = 1$), the basin selection property is preserved (Fig. 2(c),2(d)).

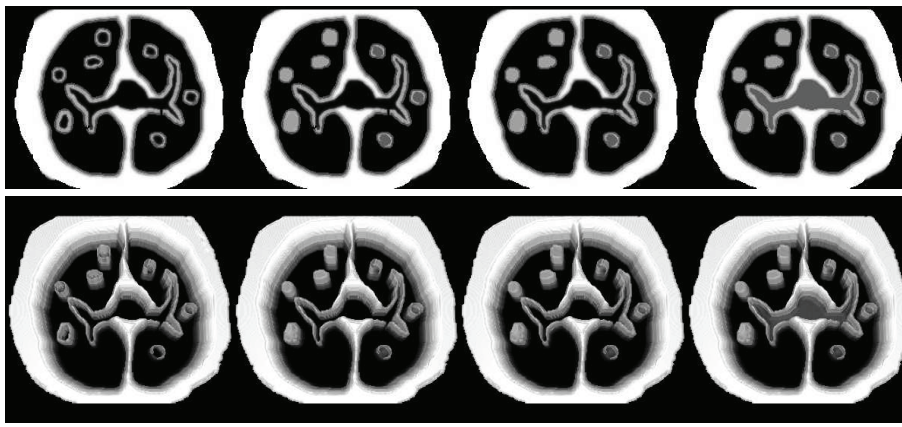


(a) f (b) FSD_f^t (c) FSD_f^T (d) $FSD_f^T - FSD_f^t$

Fig. 1. Illustration of the FSD filtering principle used in airway candidates selection in [9].

Note however that, the CT acquisition protocol may impact on the topographic properties of the image relief which describes the airway lumen as "basins" in the lung parenchyma. Consequently, several filters may be applied to the original data (smoothing, median, directional) prior to the candidate selection scheme. Their parameters will be discussed in Section 3. The airway candidates are obtained according to the procedure synthesized in Fig. 3, while Fig. 4 illustrates some intermediate results. First, the trachea cross-section is detected on the first image of the MSCT volume, then the whole trachea is extracted down to the level of the carina (bifurcation into the main bronchi), using a region-growing approach, Fig. 4(a). The maximum size of the trachea cross-section, T, is simultaneously computed. A successive filtering of the MSCT relief f using $C_m FSD_f^n$ is performed as in Fig. 2, first with $n = 3$ for noise suppression,

then using $n = T$ for airway flooding. An adaptive thresholding of the difference $C_{fill}(f) = C_m FSD_f^T - C_m FSD_f^t$ provides a first candidates set for airways, Fig. 4(b). The threshold is automatically computed on the non-zero region of $C_{fill}(f)$ which is spatially connected with the trachea. The threshold value is given by $s = \mu + 2\sigma$, where μ and σ denote respectively the mean and standard deviation estimated for $C_{fill}(f)$ values belonging to the [1, 25] interval. An artifact reduction procedure is applied slice-by-slice and validates each candidate cross-section which has at least 50% contour points adjacent to a dense tissue (bronchus wall). Parts of invalid components (Fig. 4(b), gray) which are upward/downward connected with a valid component are also preserved, Fig. 4(c).



(a) f (b) FSD_f^T (c) $C_1 FSD_f^t$ (d) $C_1 FSD_f^T$

Fig. 2. Illustration of the $C_m FSD$ filtering principle for airway candidates selection.

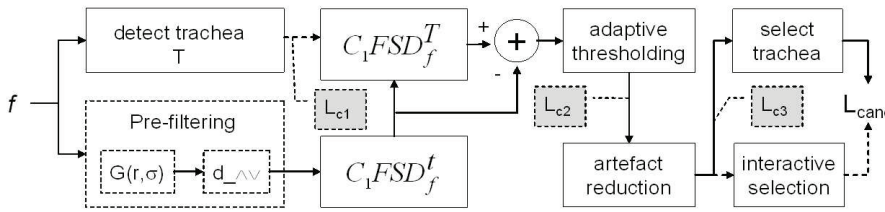
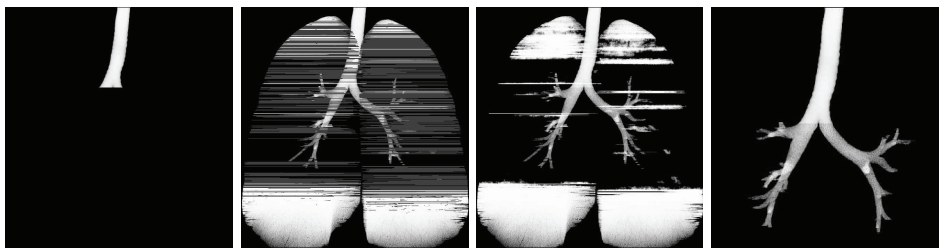


Fig. 3. Airway candidates selection: synoptic scheme.

At this point, the procedure can perform in two modes, automatic or semi-automatic. In the automatic mode, the airway candidates are selected as the only 3D component connected with the trachea, previously detected, Fig. 4(d).

In the semi-automatic mode, the user interactively selects the components to keep.



(a) L_{c1} (b) L_{c2} (c) L_{c3} (d) L_{cand}

Fig. 4. Intermediate results of the automatic airway candidate selection procedure of Fig. 3.

Note that, in particular cases where the airway lumen intensity is locally higher than the normal, due to respiratory or metallic artifacts, cardiac motion, mucoid impaction or stenoses, several connected components of airway candidates will be present, and, in this case, an interactive selection mode is recommended.

2.2 Distal airway reconstruction

The reconstruction of distal airways exploits several propagation schemes which are applied iteratively, until the convergence is reached. The propagation starts from the airway candidates which are supposed to be properly located inside the airway lumen, and progressively invests the lung regions. Three conditions have to be met by the new aggregated zones (voxels):

1. either belong to a "tunnel" configuration in terms of tissue density/image intensity, which applies mainly to the reconstruction of small-caliber airways,
2. or belong to a quasi-constant region inside the airway lumen, which applies to the segmentation of medium-caliber bronchi,
3. and avoid or, at least limit, the "leakage" in the lung parenchyma; in this respect, an additional constraint, let us call it "viscosity", can be imposed in the propagation scheme in order to locally stop the growth.

If L_{cand} denotes the airway candidates subset, $\nu \in L_{cand}$ a border point, x a free neighbor of ν (in 26-connectivity), and $f : Y \subset \Re^3 \rightarrow \Re$ the CT image intensity function, the following propagation schemes are defined in order to test the inclusion of x in L_{cand}.

A. Low-resolution tunnel propagation (LRTP)

Generally associated with a low-pass filtering (Gaussian smoothing) of the original CT data f, it requires the acceptance of at least one of the following conditions:

1. Test of tunnel configuration

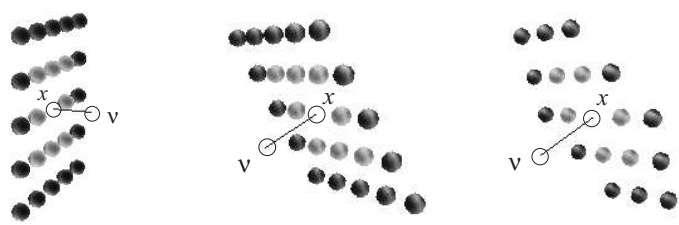
A local neighborhood of x orthogonal to the propagation direction $\overline{\nu x}$ of first ($\vartheta^{\perp}_{1,\nu}(x)$) and second order ($\vartheta^{\perp}_{2,\nu}(x)$) is used, Fig. 5. The following criteria have to be simultaneously met:

- x is a local minimum of f with respect to the neighborhood $\vartheta^{\perp}_{1,\nu}(x) \cup \vartheta^{\perp}_{2,\nu}(x)$, i.e. $\forall y \in \vartheta^{\perp}_{2,\nu}(x)$, there is no descending path on the f relief restricted to $\vartheta^{\perp}_{1,\nu}(x) \cup \vartheta^{\perp}_{2,\nu}(x)$, leading from x to y.
- if y denotes the mass center of the $\{f(\vartheta^{\perp}_{1,\nu}(x) \cup \vartheta^{\perp}_{2,\nu}(x))\}$ and $d(x,y)$ the Euclidean distance between x and y (in voxel units), then $d(x,y) \leq d_c(x)$, where $d_c(x) = 2$ for "viscous" propagation, otherwise $d_c(x) = \{\sqrt{3}, 2, 3\}$ for x $\{6-, 18-, 26-\}$ connected neighbor of ν.
- if ρ denotes the correlation coefficient computed between the "relief" of $f(\vartheta^{\perp}_{1,\nu}(x) \cup \vartheta^{\perp}_{2,\nu}(x))$ and a bowl-shape, x is added to L_{cand} if either (1) $\rho \geq 0.5$, or (2) $0.2 < \rho < 0.5$ and a f value increase in the $\overline{\nu x}$ direction is checked, $f(x + \overline{\nu x}) > min\{f(\vartheta^{\perp}_{1,\nu}(x))\}$ and $f(x + 2\overline{\nu x}) > min\{f(\vartheta^{\perp}_{2,\nu}(x))\}$.

2. Test of quasi-constant region inside the airway lumen

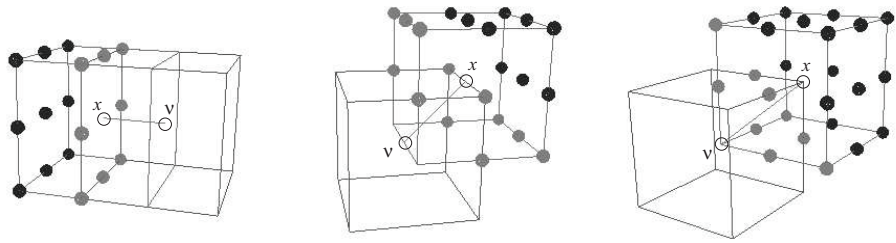
Such test is performed in a 5-voxel length region in the $\overline{\nu x}$ propagation direction.

- first, to avoid the leakage in the parenchyma, a f value large enough to be considered as airway wall should be found in a cone-shaped neigborhood of x, on each direction $\overline{\nu y}$, $y \in \vartheta_{n,\nu}(x)$ (Fig. 6), within a distance of $T/3$, where T denotes the size of the trachea cross-section (*cf.* §2.1).
- second, considering the directional neighborhoods of x in Fig. 6, the average value of f computed on each of them should satisfy the relation: $E\{f(\vartheta_{n,\nu}(x))\} \leq E\{f(\vartheta_{f,\nu}(x))\}$.
- finally, the f-value oscillation amplitude in the $\overline{\nu x}$ direction should be smaller than a constant $k_o(x) = \{7, 10, 14\}$ for x $\{6-, 18-, 26-\}$ connected neighbor of ν. The f-value oscillation amplitude is computed either along the $\overline{\nu x}$ direction, in normal propagation mode, or within the "far" neighborhood shifted in each point along the $\overline{\nu x}$ segment, $\bigcup_{y \in \overline{\nu x}} \vartheta_{f,\nu}(y)$, in "viscous" propagation mode.



(a) 6-c configuration. (b) 18-c configuration. (c) 26-c configuration.

Fig. 5. First and second order neighborhoods of x in the plane orthogonal to $\overline{\nu x}$, $\vartheta^{\perp}_{1,\nu}(x)$ (light grey) and $\vartheta^{\perp}_{2,\nu}(x)$ (dark grey), respectively, according to the $\nu - x$ connectivity.



(a) 6-c configuration. (b) 18-c configuration. (c) 26-c configuration.

Fig. 6. Directional "near" and "far" neighborhoods of x with respect to its position relative to ν (6-, 18-, 26- connectivity): $\vartheta_{n,\nu}(x)$ (grey) and $\vartheta_{f,\nu}(x)$ (black).

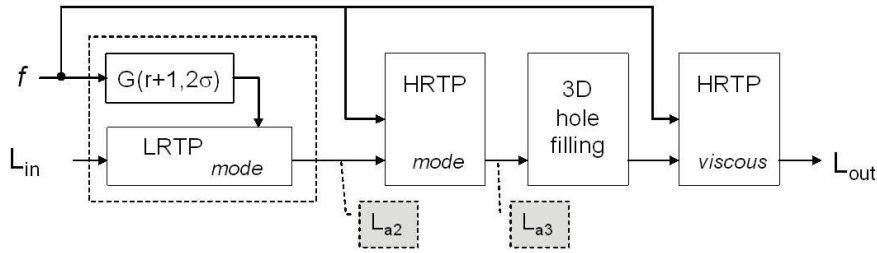
B. High-resolution tunnel propagation (HRTP)

The HRTP scheme includes the same conditions as the LRTP and adds a regularizing one which may facilitate the acceptance of a ν neighbor, x, in the airway lumen set L_{cand}.

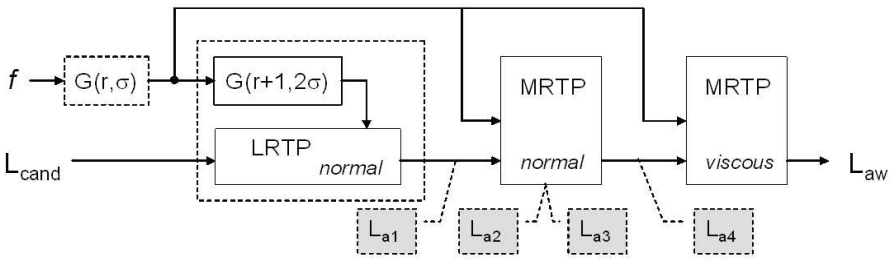
3. Regularization condition

In addition to the first two requirements of the *tunnel configuration test* in LRTP, x has to fulfill a neighborhood occupancy condition: $N_o(x) + \mu_f(x) > N_f(x) + \mu_o(x) + \sigma_o(x) + T_{100}(f(x))$, where $N_o(x)$ and $N_f(x)$ denote respectively the number of occupied and free 26-neighbors of x, $\mu_o(x)$ and $\mu_f(x)$ the average f values computed on $N_o(x)$ and $N_f(x)$, $\sigma_o(x)$ the standard deviation of f on $N_o(x)$, and T_a the thresholding operator, $T_a(u) = \{u,\ if\ u \geq a;\ 0\ otherwise\}$.

The distal airway reconstruction scheme implemented is summarized in Fig. 7. Note that, as in § 2.1, a prefiltering of the original data f might be needed according to the CT scanning protocol used. These parameters will be discussed in Section 3. A multi-resolution tunnel propagation (MRTP) module is designed as a succession of a LRTP and a HRTP (which can be applied either in normal or "viscous" mode), followed by a 3D hole filling and by a HRTP in "viscous" mode, Fig. 7(a). The distal airway reconstruction starts by reinforcing the airway candidates by means of a LRTP in normal mode, Fig. 7(b). Then, the MRTP scheme is applied successively in normal and "viscous" mode. The intermediate propagation results are illustrated in Fig. 8.

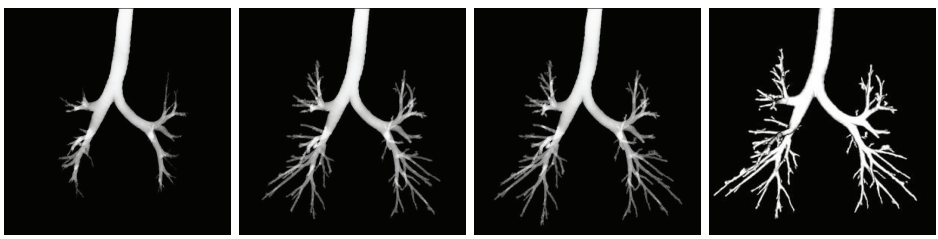


(a) Multi-resolution tunnel propagation (MRTP) scheme.



(b) Distal airways reconstruction.

Fig. 7. Reconstruction of distal airways: synoptic scheme. Note that, if no data prefiltering is used ($r = 0$), then $\sigma = 0.1$ in the LRTP scheme.



(a) L_{a2} (b) L_{a3} (c) L_{a4} (d) L_{aw}

Fig. 8. Intermediate and final results of the distal airway segmentation procedure of Fig. 7(b).

3 Results and discussion

The developed segmentation approach was assessed in the framework of the EXACT'09 challenge (http://image.diku.dk/exact/) by a group of independent experts[11]. The evaluation database included 20 pairs of anonymized MSCT cases of various pathologies, acquired in different hospital centers using several CT scanners and protocols (Table 1). Half of them represented the training test, for algorithm parameter tune-up, and the other half constituted the testing set.

A ground truth was defined as the union of all valid airway segments from all submitted segmentations, and all submissions were subsequently evaluated with respect to this ground truth.

Table 1. Description of the evaluation dataset, the algorithm parameters used and the recorded running time (test set only). Here, G denotes a Gaussian filter of parameters G(1,0.2).

CASE number	MSCT scanner	Reconstruction kernel	X-ray dose (mAs)	Collimation (mm)	z sampling (mm)	Prefiltering Fig. 3 / Fig. 7(b)	Running time (min:sec)
1, 21	Siemens	B50f	200	0.6	0.5	G / G	6:07
2, 22	Siemens	B50f	200	0.6	0.6	G / G	14:11
3, 23	Siemens	B50f	200	0.75	0.5	G / G	12:26
4, 24	Toshiba	FC12	10	1.0	0.8		7:33
5, 25	Toshiba	FC10	150	1.0	0.8		7:34
6, 26	Toshiba	FC12	10	1.0	0.8		3:20
7, 27	Toshiba	FC10	150	1.0	0.8		2:45
8, 28	Siemens	B30f	300	1.25	?		3:59
9, 29	Siemens	B50f	300	1.25	?		4:17
10, 30	Philips	D	120	1.0	1.0		2:04
11, 31	Philips	D	120	1.0	1.0		2:23
12, 32	Philips	D	120	1.0	1.0		3:31
13, 33	Siemens	B60f	222	1.0	1.0		5:20
14, 34	Siemens	B60f	321	1.0	0.8		5:23
15, 35	GE	STD	357	0.625	0.625		4:53
16, 36	Philips	C	184	1.0	0.7		6:57
17, 37	Philips	B	64	1.0	0.7		5:46
18, 38	Philips	C	51	1.0	0.7		5:40
19, 39	Siemens	B70f	361	1.0	1.0	G / G	4:29
20, 40	Siemens	B70s	108	1.0	1.0	$G + d^3_{f\wedge\vee}$ / G	3:56

Airway segments were visually assessed on a set of extracted slices from both a reoriented view and a reformatted view with straightened airway centerlines. Each segment was scored as "correct" or "wrong", by at least two observers. The criterion used is whether the extracted airway segment indeed belongs to the airway tree; the exact airway shape and dimensions were not taken into account.

In our case, the only parameter tuning was at the level of prefiltering of the original data (Fig. 3, 7(b)). The filters considered here were: a Gaussian smoothing $G(r, \sigma)$, where r denotes the kernel radius and σ the considered standard deviation, and/or a directional filter $d^n_{f\wedge\vee}(x)$ which returns the minimum value among the maximum values of f computed along each of the 13 main directions, in a length interval [-n, n]. The use of prefiltering is decided during a protocol-learning phase applied only on the training set, and consisting of three runs of the algorithm, first without filtering, second using $G(r, \sigma)$, and third, using $G(r, \sigma)$ and adding $d^n_{f\wedge\vee}(x)$ in the candidate selection scheme (Fig. 3)

only. The parameters providing the best average results are selected according to the CT manufacturer, reconstruction kernel applied and the X-ray dose ($\leq$ or $>$ 200 mAs). According to the protocol database thus obtained, the prefiltering parameters are automatically selected for the cases in the test set. This ensures the automatism of the proposed approach. For the cases where the prefiltering was selected, the parameters are mentioned in Table 1. Note that, in the absence of the Gaussian smoothing, the low-pass filter associated with the LRTP scheme (Fig. 3, 7(b)) has the parameters G(1,0.2). The airway segmentation procedure was run in automatic mode for all 40 cases.

The evaluation results for the test dataset are shown in Table 2. The following measurements were computed by the external evaluation team and used for comparing the submitted results:

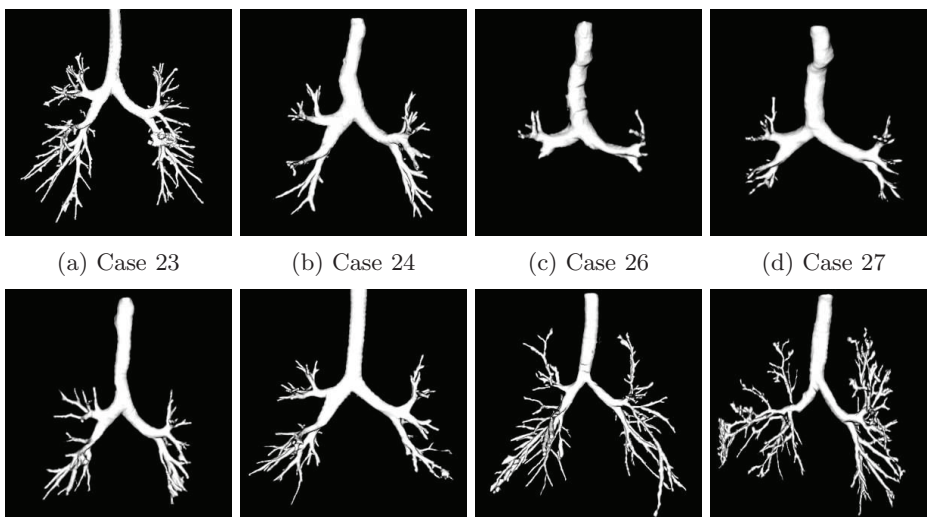
Table 2. Evaluation measures for the twenty cases in the test set.

	Branch count	Branch detected (%)	Tree length (cm)	Tree length detected (%)	Leakage count	Leakage volume (mm^3)	False positive rate (%)
CASE21	115	57.8	62.5	56.6	2	1.5	0.02
CASE22	256	66.1	193.1	58.4	34	915.6	3.02
CASE23	247	87.0	217.7	83.7	30	3973.7	11.62
CASE24	130	69.9	101.5	62.4	3	3.7	0.01
CASE25	146	62.4	115.3	45.8	0	0.0	0.00
CASE26	28	35.0	21.5	32.7	4	5.7	0.11
CASE27	83	82.2	60.3	74.4	2	2.7	0.02
CASE28	82	66.7	61.5	56.1	10	566.1	4.56
CASE29	124	67.4	80.1	58.0	11	280.4	2.02
CASE30	118	60.5	81.4	53.2	2	45.8	0.37
CASE31	117	54.7	86.7	49.4	9	280.8	1.32
CASE32	122	52.4	93.4	42.9	10	1336.1	4.97
CASE33	113	67.3	82.3	56.0	3	10.8	0.11
CASE34	282	61.6	201.6	56.4	15	68.1	0.17
CASE35	192	55.8	134.6	43.5	8	659.8	2.54
CASE36	239	65.7	258.2	62.6	13	88.7	0.26
CASE37	104	56.2	82.5	46.4	1	71.7	0.35
CASE38	52	53.1	40.1	60.4	1	0.4	$<$0.01
CASE39	336	64.6	256.3	62.6	38	1241.8	3.45
CASE40	269	69.2	217.8	56.3	44	1716.4	4.27
Mean	157.8	62.8	122.4	55.9	12.0	563.5	1.96
Std. dev.	85.1	11.0	73.8	11.2	13.5	959.3	2.86
Min	28	35.0	21.5	32.7	0	0.0	0.00
1st quartile	104	55.8	62.5	46.4	2	3.7	0.02
Median	123	63.5	90.0	56.3	9	80.2	0.36
3rd quartile	256	69.2	217.7	62.6	30	1241.8	4.27
Max	336	87.0	258.2	83.7	44	3973.7	11.62

1. Branch count: the number of branches that are detected correctly. A branch is considered detected as long as the length of the centerlines is more

than 1 mm; **2. Branch detected:** the fraction of branches that are detected, with respect to the branches present in the ground truth; **3. Tree length:** the sum of the length of the centerlines of all correctly detected branches; **4. Tree length detected:** the fraction of tree length in the ground truth that is detected correctly; **5. Leakage count:** the number of unconnected groups of "correct" regions that are neighboring with a "wrong" region (indicates how easy/difficult it is to manually separate leakages from the correctly detected branches); **6. Leakage volume:** the volume of regions that are wrongly detected; **7. False positive rate:** the fraction of the volume of regions that are detected wrongly over the volume of all detected regions.

Fig. 9 illustrates some representative results. Fig. 8(d) concerns the case 22. Note that the airway segmentation accuracy strongly depends on the radiation dose, reconstruction kernel, lung inflation volume, and on the type/severity of the pathology. The presence of image artifacts (respiratory, cardiac motion, metallic) has generally a negative impact on the reconstruction result, mainly when the automatic mode is selected, as airway candidates disconnected from the trachea will be lost and might not be retrieved in the propagation phase. When working in semi-automatic mode, higher accuracy may be achieved at the expense of interaction time.



(a) Case 23 (b) Case 24 (c) Case 26 (d) Case 27

(e) Case 32 (f) Case 35 (g) Case 36 (h) Case 39

Fig. 9. Some results of the developed airway segmentation procedure. Illustration using opaque volume rendering.

The algorithm was run on a Dell PowerEdge workstation equipped with Linux Red Hat 5.3 operating system, Intel Xeon CPU 3.4 GHz and 6 GBytes RAM

memory. The execution time recorded for each case of the test set is mentioned in Table 1.

At this stage, we cannot perform a more in-depth analysis of the results obtained with respect to the acquisition protocols used, as some information is missing both in terms of type of pathology and CT protocol parameters that would be able to explain differences shown in airway segmentation.

Concerning the participation to the EXACT'09 airway segmentation challenge, the developed approach reaches a high grading among the fifteen participating teams, as revealed by the average scores published on the EXACT'09 webpage, http://image.diku.dk/exact/results.php. This study stresses out the importance of the MSCT acquisition protocol set-up according to the type of the planned investigation, irrespective to the 3D segmentation approach used.

References

1. Mori, K., Hasegawa, J., Toriwaki, J., Anno, H., Katada, K.: Recognition of Bronchus in Three-Dimensional X-ray CT Images with Applications to Virtualized Bronchoscopy System. In: Proceedings of the 13th International Conference on Pattern Recognition. Volume 3. (1996) 528–532
2. Park, W., Hoffman, E.A., Sonka, M.: Fuzzy logic approach to extraction of intrathoracic airway trees from three-dimensional CT images. In: Proc SPIE. Volume 2710. (1996) 210–219
3. Sonka, M., Park, W., Hoffman, E.A.: Rule-based detection of intrathoracic airway trees. IEEE Transactions on Medical Imaging **15**(3) (1996) 314–326
4. Reinhardt, J.M., D'Souza, N.D., Hoffman, E.A.: Accurate measurement of intrathoracic airways. IEEE transactions on medical imaging **16**(6) (1997) 820 – 827
5. Kiraly, A.P., Higgins, W.E., McLennan, G., Hoffman, E.A., Reinhardt, J.M.: Three-dimensional human airway segmentation methods for clinical virtual bronchoscopy. Academic Radiology **9**(10) (2002) 1153–1168
6. Schlatholter, T., Lorenz, C., Carlsen, I.C., Renisch, S., Deschamps, T.: Simultaneous segmentation and tree reconstruction of the airways for virtual bronchoscopy. In: Proc SPIE. Volume 4684. (2002) 103–113
7. Tschirren, J., Hoffman, E.A., McLennan, G., Sonka, M.: Intrathoracic airway trees: Segmentation and airway morphology analysis from low-dose CT scans. IEEE Transactions on Medical Imaging **24**(12) (December 2005) 1529 – 1539
8. Fetita, C., Prêteux, F., Grenier, P.: 3D reconstruction of the bronchial tree in volumetric CT: application to ct bronchography. Journal of Electronic Imaging **15**(2) (October 2006) 023004–1:17
9. Fetita, C., Prêteux, F., Beigelman-Aubry, C., Grenier, P.: Pulmonary airways: 3D reconstruction from multi-slice CT and clinical investigation. IEEE Transactions on Medical Imaging **23**(11) (November 2004) 1353–1364
10. Vincent, L.: Morphological gray scale reconstruction in image analysis: applications and efficient algorithms. IEEE Trans on Image Processing **2**(2) (1993) 176–201
11. Lo, P., van Ginneken, B., Reinhardt, J., de Bruijne, M.: Extraction of airways from CT (EXACT'09). In: Second International Workshop on Pulmonary Image Analysis. (2009)

Airway Segmentation Framework for Clinical Environments

Juerg Tschirren[1], Tarunashree Yavarna[2], and Joseph M. Reinhardt[2]

[1] VIDA Diagnostics, Inc.
100 Oakdale Campus, 225 TIC
Iowa City, IA 52242, USA,
juerg@vidadiagnostics.com,
http://www.vidadiagnostics.com
[2] The University of Iowa, Dept. of Biomedical Engineering,
Iowa City, IA 52242, USA

Abstract. A segmentation framework for the identification of human airway trees in high resolution computed tomography (CT) images is presented. This framework consists of a fully automated segmentation algorithm, supplemented by software editing tools that allow the user to correct the airway segmentation result if needed. The algorithm and tools presented in this paper have been successfully applied on more than 10,000 CT scans.

1 Introduction

The airway segmentation algorithm and manual airway editing tools presented here are part of VIDA Diagnostic's "Pulmonary Workstation 2.0" (PW2), a commercial software suite for the analysis of human lung CT scans. The software requirements for a commercial product such as PW2 differ substantially from those used in academic settings. The primary goal is to run a segmentation task in as little time as possible, and have it return an acceptable result in the majority of cases. An algorithm that finished in a few seconds and returns a usable result is strongly preferred over an algorithm with a run time of several minutes or longer, even if this slower algorithm returns significantly better results.

Beyond the automated methods it is important to have manual editing tools available to correct results as needed. No algorithm can guarantee 100% perfect results in all possible cases and the user needs to be able to correct segmentation results as needed, using efficient and intuitive tools.

In this paper we present an airway segmentation framework that consists of two major parts: 1) a fully automated airway segmentation algorithm, and 2) a set of software tools that allow the user to edit the airway segmentation result if needed.

The aim of the airway segmentation algorithm presented here is to identify the whereabouts of the airway lumen. The algorithm presented here does *not*

Juerg Tschirren and Joseph Reinhardt are shareholders in VIDA Diagnostics, Inc.

identify the exact location of the airway wall (as for example used to determine exact airway diameters). Exact inner and outer airway wall localization is done by the airway measurement algorithm, which we described elsewhere as it lies outside the scope of this paper.

2 Methods

2.1 Automated Airway Segmentation

The automated airway segmentation algorithm is split into two parts: 1) finding a seed point within the trachea, and 2) growing the airway tree. They are described separately below.

Trachea Finder The trachea seed point required to grow the airway tree is found using the following steps:

1. *Threshold axial slices.* Every axial slice s_a in the slice range $[15, \frac{N_{\text{axial}}}{2}]$ of the CT volume is transformed into a thresholded axial image $\mathcal{T}_a$ using

$$\mathcal{T}_a(x,y) = \begin{cases} 255 & \text{if } |F(x,y)| > |B(x,y)| \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

with the foreground and background pixel sets defined as

$$\begin{aligned} F(x,y) &= \{(x',y') : \forall (x',y') \in N_9(x,y), d(s_a(x',y')) \leq t\} \\ B(x,y) &= \{(x',y') : \forall (x',y') \in N_9(x,y), d(s_a(x',y')) > t\}, \end{aligned} \tag{2}$$

respectively. $N_9(x,y)$ represents the 3×3 neighborhood at (x,y), inclusive (x,y), and $d(\cdot)$ returns the CT density of a voxel in s_a, measured in Hounsfield Units (HU).

2. *Find & classify foreground objects.* In every thresholded slice t_a a connected component analysis [8] is performed and the border polygon of every foreground object is found using the *boundary tracing* algorithm described in [8, p. 142]. The compactness $\sigma \in [0,1]$ is computed for every object with

$$\sigma = \frac{4\pi A}{c^2} \tag{3}$$

with $\sigma = 1$ in case of a perfectly circular object, and the value of σ getting smaller the less circular the object. The area A is computed based on the number of object pixels and the circumference c is computed by adding up the lengths of the segments along the border polygon.
Only objects with $\sigma > 0.2$ and $40 \text{ mm}^2 < A < 500 \text{ mm}^2$ are used for further processing.

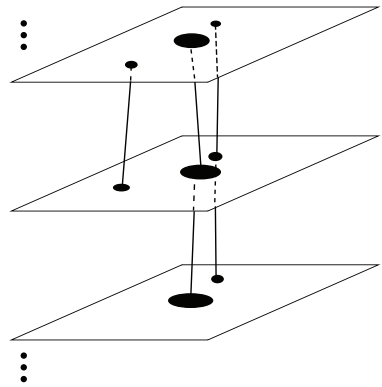


Fig. 1. Chains of foreground objects between individual axial slices are built by connecting elements that are close enough and similar enough to each other.

3. *Build foreground object chains.* The similarity ς between two foreground objects f_1 and f_2 is defined as

$$\varsigma(f_1, f_2) = 1 - \frac{d(f_1, f_2)}{r_{\text{circ f1}}} \tag{4}$$

with the Euclidean distance d [voxels] between centroids

$$d(f_1, f_2) = \sqrt{(x_{\text{centroid f1}} - x_{\text{centroid f2}})^2 - (y_{\text{centroid f1}} - y_{\text{centroid f2}})^2} \tag{5}$$

and the circle-equivalent radius

$$r_{\text{circ f1}} = \sqrt{\frac{A_{f1}}{\pi}} \tag{6}$$

Only object pairs with $d(f_1, f_2) \leq 2$ are considered.
Given the set $\mathcal{F}$ all foreground objects, find the object pairs (f_1, f_2) with the highest similarity and arrange them into one or more object chain(s) c (Figure 1), stored in the set $\mathcal{C}$ ($c \in \mathcal{C}$) with

for all $f_1 \in \mathcal{F}$ **do**
  $s_{\max} \leftarrow 0$
  $f_{\max} \leftarrow$ null
  for all $f_2 \in \mathcal{F}, f_1 \neq f_2$ **do**
    $s \leftarrow \varsigma(f_1, f_2)$
    if $s > s_{\max}$ **then**
      $s_{\max} \leftarrow s$
      $f_{\max} \leftarrow f_2$
    end if
  end for
  if $s_{\max} > 0.6$ **then**
    $\mathcal{C} = \mathcal{C} \cup (f_1, f_{\max})$

end if
end for

4. *Find trachea chain.* For each $c \in \mathcal{C}$, compute the average radius r_{avg} with

$$r_{\text{avg}} = \frac{1}{N} \sum_{i=0}^{N-1} r_i \tag{7}$$

where N is the number of foreground objects in c, and the average variation from the volume center along the x-axis, x_{var}, defined as

$$x_{\text{var}} = \frac{1}{N} \sum_{i=0}^{N-1} \left| c_x - \frac{X}{2} \right| \tag{8}$$

where X is the volume size along the x-axis (number of voxels).
If r_{avg} and x_{var} apply to the same chain, label this chain as trachea. Otherwise compute the mean y value for all centroids along the chain

$$\bar{y}_{\text{centroid}} = \frac{1}{N} \sum_{i=0}^{N-1} c_y \tag{9}$$

for each of these two chains and label the one chain with the smaller $\bar{y}_{\text{centroid}}$ value as trachea. This is to prevent the esophagus from being labeled as trachea (assuming scan orientation such that the y-axis increases from anterior to posterior).

Growing Airway Tree Many different airway segmentation algorithms have been presented in the past. [1, 3, 10, 5, 2, 4] represent just a few of them. The core algorithm for the automatic airway segmentation is similar to the one presented by Mori et al. [1]. A queue-based breadth-first flood fill algorithm [6] is used to grow the airway lumen (Algorithm 1). This algorithm is executed iteratively by Algorithm 2, continuously increasing the threshold value for the region grow. At each iteration the number of voxels grown by the flood fill algorithm is recorded. A sudden big increase between two consecutive iteration steps n and $n + 1$ is considered a *leak*. We use the term "leak" for the local effect where grown airway lumen "mushrooms" into the surrounding lung parenchyma. This may happen due to the similar x-ray density values of lung tissue, compared with the airway lumen, and the relatively thin airway wall, particularly in small peripheral airways. Figure 2 shows an un-edited segmentation of an airway tree.

2.2 Manual Editing Tools

Deleting Airway Leaks Airway leaks can easily be removed with a few mouse clicks. First the airway tree is skeletonized using sequential 3D thinning as described in [7]. This gives us information about the topology of the tree and allows us to identify the individual segments and make them selectable in a 3D surface

Algorithm 1 Flood fill algorithm.

Require: *seed* {Coordinate of seed voxel}
Require: t {Threshold value}
  initial-color ← color(*seed*)
  color(*seed*) ← final-color
  $n \leftarrow 1$
  enqueue(Q, *seed*)
  while Q not empty **do**
    $h \leftarrow$ head(Q)
    dequeue(Q)
    for each $n \in$ neighbors(h, 6-neighborhood) **do**
      if color(n) = initial-color and density(n) $< t$ **then**
        color(n) ← final-color
        $n \leftarrow n + 1$
        enqueue(Q, n)
      end if
    end for
  end while
  return n

Algorithm 2 Iterative Growing. The maximal grow rate is set to $g_{\max} = 1.6$ and the maximum voxel count is set to $n_{\text{voxels max}} = 500,000$

Require: *seed* {Coordinate of seed voxel}
  $t \leftarrow$ density(*seed*)
  $n_{\text{voxels}} \leftarrow$ flood_fill(c_{seed}, t)
  repeat
    $t \leftarrow t + 1$
    $n_{\text{voxels prev}} \leftarrow n_{\text{voxels}}$
    $n_{\text{voxels}} \leftarrow$ flood_fill(c_{seed}, t)
    $g \leftarrow \frac{n_{\text{voxels}}}{n_{\text{voxels prev}}}$
    $i \leftarrow i + 1$
    erase current segmentation result
  until $g > g_{\max}$ or $n_{\text{voxels}} > n_{\text{voxels max}}$
  flood_fill(c_{seed}, $t - 1$)

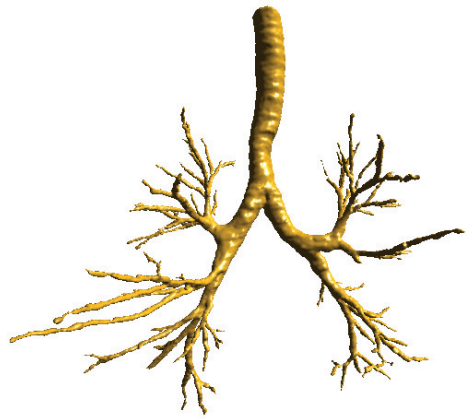
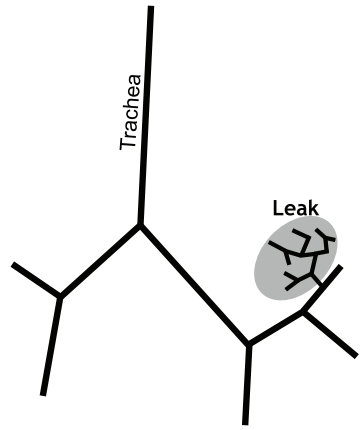


Fig. 2. Segmentation result of a CT scan at total lung capacity (TLC). 281 individual branches. No manual corrections. Segmentation time 17.5 s on an Intel Xeon^{TM} CPU based machine running at a clock speed of 2.33 GHz. CT scan from VIDA's test database (not a subject from the EXACT09 contest database).

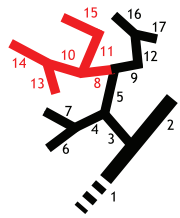
rendering. The user then selects an arbitrary segment within the leak and uses the mouse wheel to progress the selection up and down the tree as illustrated in the graph view in Figure 3. Turning the mouse wheel in one direction increases the number of selected segments, and turning the mouse wheel in the other direction de-selects the just selected segments. This allows the user to iteratively find the selection size that includes as many leak segments as possible without selecting actual airway segments. At any time the currently selected segments can be deleted by right-clicking on the selection. Figure 4 shows the actual progress in a sequence of screen shots taken in PW2.

Adding New Sub-Trees The user can add a missing sub-tree by placing a seed point inside the lumen, near the end of the current segmentation result. Figure 5(a) shows a screen shot from PW2 illustrating this. The user can select between three different threshold levels, which determines the growth rate. The computer then attempts to add the missing branch(es) starting at this seed point, using the algorithm described in Section 2.1 above.

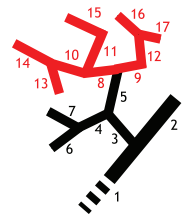
The region-grow algorithm alone can not always connect the newly added sub-tree back to the main tree. This is for example the case when a radio-dense object such as a mucus plug blocks the airway lumen. If connectivity is not automatically established we use Dijkstra's algorithm [9], starting from the user-place seed point, to find the shortest path back to the main tree. The cost function used favors low densities, and a penalty is added which progressively increases for longer paths.



(a) Leak in left upper lobe.

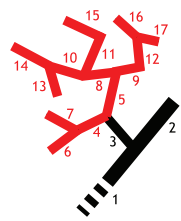


(b) User selects segment 8 (s8). All segments topologically underneath s8 are automatically selected.

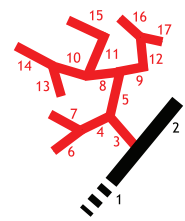


(c) The first mouse-wheel increment expands the selection to the sibling of s8 (s9) and all segments topologically underneath it.

(d) The second mouse-wheel increment selects s8's parent segment s5.

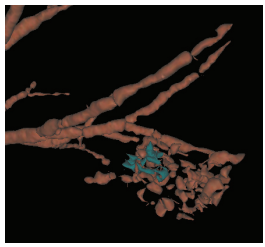


(e) The third mouse-wheel increment expands the selection to s5's sibling and the segments topologically underneath it.

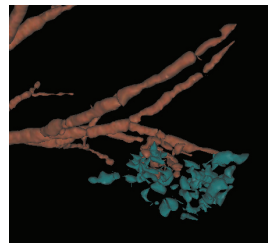


(f) After fourth and last mouse-wheel increment, s5's parent segment s3 is selected.

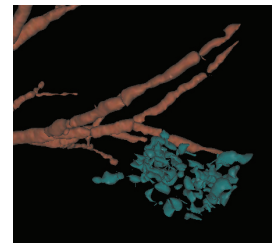
Fig. 3. Leak selection in graph view. (a): overall view. (b)–(f): detail view of leak with red/gray representing individual user selection steps. See main text for detailed explanation.



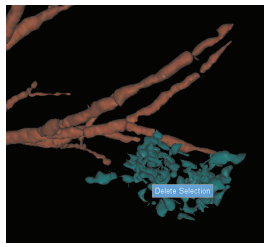
(a) User selected an arbitrary segment within a leak.



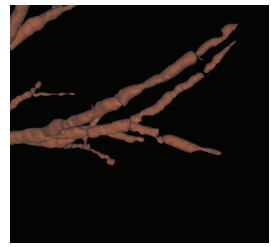
(b) After first mouse-wheel increment.



(c) After second mouse-wheel increment.

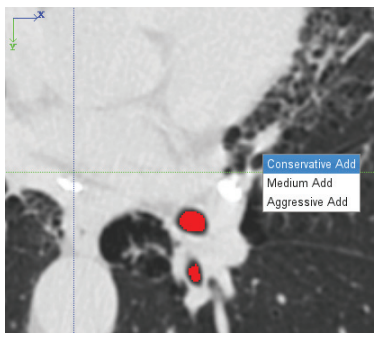


(d) After third mouse-wheel increment, all leak segments are selected. User right-clicks on selection and confirms leak removal through pop-up menu.

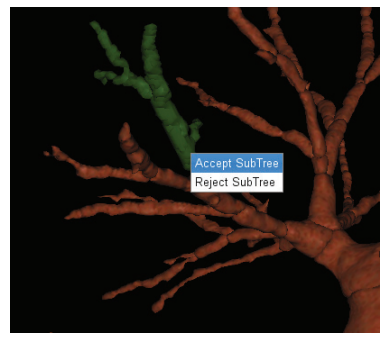


(e) Leak removed.

Fig. 4. Leak removal in Pulmonary Workstation 2.0. A leak can easily be removed with a few mouse clicks.



(a) User placed a seed point in CT view and chooses growth rate from pop-up menu.



(b) User confirms newly grown tree in 3D surface rendering through mouse click in pop-up menu.

Fig. 5. Manual growing of new sub-tree.

The added sub-tree is shown highlighted both in the CT viewer of PW2 as well as in a 3D surface rendering of the complete airway tree. The user can then examine the new addition and either accept or reject it with a simple mouse click [Figure 5(b)].

3 Validation

Twenty CT scans (identified as CASE21–CASE40 below) where made available by the organizers of the *Extraction of Airways from CT 2009 (EXACT09)* contest at the 2^{nd} *International Workshop on Pulmonary Image Analysis*, MICCAI 2009, London. In a first step each case was segmented automatically by PW2. The standard built-in airway segmentation algorithm was used using the default parameters. Subsequently every case was examined by a human operator. Using the PW2 manual editing tools leaks were removed and sub-branches were added if applicable. The final segmentation results were sent to the EXACT09 organizers for evaluation where they were compared against a gold standard. This gold standard was not made available to the contestants prior to scoring.

4 Results

Table 1 was received from the EXACT09 organizers. The last column in this table, labeled "Manual correction time [min]", was added by the authors of this paper, based on logs kept during the segmentation task.

5 Discussion

Table 1 shows that the method presented here achieved a relatively high number of identified airway branches and a low leakage rate. However, manually growing a lot of additional airway branches throughout the airway tree comes at a price, as can be seen in the last column of Table 1. The operator spent on average almost an hour per case trying to retrieve as many branches as possible.

The PW2 software was optimized for use with a specific imaging protocol. This protocol is a based on a diagnostic-quality clinical chest CT. Deviations from the prescribed protocol in dose, reconstruction filter, slice collimation, pitch, lung volume, and other critical imaging parameters may reduce image quality and segmentation performance. The EXACT test cases span a wide range of image acquisition settings, and arguably, most cases are not high-quality diagnostic images. However, even for cases where the PW2 automatic segmentation produced a suboptimal segmentation, the flexible manual editing tools allowed the operator to add to and refine the segmentation if desired.

But from our observations in both clinical applications as well as in research studies where PW2 is used today we know the actual operator time spent per tree is, in most cases, significantly lower. This is because the automatically identified airway tree is often sufficient for the purpose of the specific analysis. And if more

Table 1. Evaluation measures for the twenty cases in the test set.

	Branch count	Branch detected (%)	Tree length (cm)	Tree length detected (%)	Leakage count	Leakage volume (mm^3)	False positive rate (%)	Manual correction time [min]
CASE21	116	58.3	69.6	63.0	0	0.0	0.00	10
CASE22	218	56.3	181.2	54.8	17	125.3	0.88	180
CASE23	159	56.0	110.6	42.5	22	141.8	1.09	90
CASE24	113	60.8	89.0	54.7	6	275.1	1.58	30
CASE25	169	72.2	145.7	57.8	13	214.6	1.08	50
CASE26	24	30.0	17.6	26.8	0	0.0	0.00	0
CASE27	89	88.1	71.1	87.8	15	265.1	3.03	70
CASE28	96	78.0	72.6	66.3	7	62.7	0.83	50
CASE29	129	70.1	91.8	66.5	10	207.8	2.20	90
CASE30	142	72.8	102.9	67.4	7	81.2	0.91	25
CASE31	147	68.7	111.2	63.3	5	76.3	0.62	50
CASE32	156	67.0	137.3	63.0	14	285.0	1.80	80
CASE33	138	82.1	118.5	80.6	9	79.3	1.00	105
CASE34	329	71.8	256.5	71.7	20	301.8	1.10	70
CASE35	234	68.0	177.3	57.3	18	67.8	0.41	60
CASE36	105	28.8	101.4	24.6	0	0.0	0.00	0
CASE37	138	74.6	127.2	71.6	10	121.8	0.83	90
CASE38	61	62.2	44.8	67.4	5	164.5	2.15	35
CASE39	135	26.0	113.4	27.7	5	188.0	1.86	0
CASE40	275	70.7	243.6	63.0	26	517.2	2.34	90
Mean	148.7	63.1	119.2	58.9	10.4	158.8	1.19	58.75
Std. dev.	71.0	17.1	59.8	16.9	7.5	128.0	0.84	44.10
Min	24	26.0	17.6	24.6	0	0.0	0.00	0.00
1st quartile	105	56.3	72.6	54.7	5	67.8	0.62	28.75
Median	138	68.4	110.9	63.0	10	133.5	1.04	55.00
3rd quartile	218	74.6	177.3	71.6	18	275.1	2.15	90.00
Max	329	88.1	256.5	87.8	26	517.2	3.03	180.00

peripheral airways need to be identified then these are often restricted to a few pre-defined paths. Therefore most small-sized airways need not be delineated. For the test data in the EXACT challenge, the operator was instructed to add all visible airway branches. This is generally not necessary or desirable for a clinical evaluation of the airways. For procedure planning purposes such as a biopsy or the placing of a stent normally only one single airway path is required. Clinical trials often only look at three or four sentinel airway paths.

In that sense the relative high operator times reported here are not representative for most clinical applications of the PW2 software and could be called a "contest artifact". In the day to day PW2 operation most cases are analyzed much faster.

6 Conclusion

The airway segmentation algorithm presented here is part of VIDA Diagnostic's "Pulmonary Workstation 2.0" (PW2) software suite, an FDA 510(k) approved software package for the analysis of CT scans of human lungs. PW2 is used in clinical settings for tasks such as the quantification of lung disease and bronchoscopic procedure planning. Today PW2 is also utilized as the central software tool in various clinical trials to determine the efficacy of new treatment methods, and in research studies that seek correlations between lung disease and specific genotypes and phenotypes.

So far over 10,000 CT datasets have been analyzed with the methods described here, and several dozen new cases are added every day.

References

1. Mori, K. and Suenaga, Y. and Toriwaki, J. Automated anatomical labeling of the bronchial branch and its application to the virtual bronchoscopy. IEEE Transactions on Medical Imaging. Vol. 19. 103–114 (2000).
2. van Ginneken, B. et al. Robust Segmentation and Anatomical Labeling of the Airway Tree from Thoracic CT Scans. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008. Part 1. 219–226. (2008)
3. Schlathölter, T. et al. Simultaneous segmentation and tree reconstruction of the airways for virtual bronchoscopy. Proc. SPIE, vol. 4684. 103–113. (2002)
4. Lo, P. et al. Voxel classification-based airway tree segmentation. Proc. SPIE, vol. 6914 (2008)
5. Tschirren, J. et al. Intrathoracic airway trees: segmentation and airway morphology analysis from low-dose CT scans. IEEE Trans. Med. Im. 24(12), 1529–1539. 2005
6. Silvela, J. and Portillo, J. Breadth-first search and its application image processing problems. IEEE Transactions on Image Processing. Vol. 10, Issue 8. 1194–1199 (2001).
7. Palágyi, K. and Tschirren, J. and Sonka, M. Quantitative analysis of intrathoracic airway trees: methods and validation. In Proc. 18th Int. Conf. Information Processing in Medical Imaging, IPMI 2003, Ambleside, UK, Lecture Notes in Computer Science 2732, Springer, 222–233, 2003.

8. Sonka et al. Image Processing, Analysis, and Machine Vision. 2nd ed. Brooks/Cole Publishing. (1998).
9. Cormen, T. et al. Introduction to Algorithms. McGraw-Hill (2003).
10. Kiraly, Atilla and Higgins, William E., and McLennan, Geoffrey and Hoffman, Eric A. and Reinhardt, Joseph M. 3D human airway segmentation for clinical virtual bronchoscopy. Academic Radiology. Vol. 9, Number 10. 1153–1168 (2002).

Three-Step Segmentation of the Lower Airways with Advanced Leakage-Control

Silvia Born[1], Dirk Iwamaru[2], Matthias Pfeifle[3], and Dirk Bartz[1]

[1] Visual Computing, ICCAS, Universität Leipzig, Germany
[2] CADMEI GmbH, Ingelheim, Germany
[3] Department of Neurosurgery, University Hospital Tübingen, Germany

Abstract. Segmentation of the tracheo-bronchial tree of the lung serves as an important tool for diagnosis and treatment planning of various pathologies, e.g. by allowing accurate volume measurements or detecting malformations. However, segmenting the airways of the human lung is notoriously difficult. This is due to the small size of the participating anatomical structures, which are moreover subject to partial volume and noise effects. Limited intensity contrasts between air and lung parenchyma also complicate segmentation. In this paper, we present our hybrid segmentation method consisting of three main steps, which are iterated until a satisfactory result is achieved. User interaction is limited to the specification of a seed point inside the easily detectable trachea. Further, we discuss the performance of our method within the EXACT09 challenge, where 20 test datasets with varying quality and pathologies had to be processed.

1 Introduction

Several pathologies can jeopardize a sufficient lung function. Among them are tumors, pulmonary embolism, collapse of the lungs (atelectasis), pneumonia, emphysema, asthma, and many more. For a proper diagnosis and treatment, the respective pathologies need to be identified and in some cases quantified. In the case of lung-surgery (i.e., for tumor treatment), this information is necessary for the intervention planning where the anatomical relation of diseased bronchi to non-diseased areas is required pre-operatively, i.e. to provide a safe distance to essential structures and to determine resectability.

The standard imaging method to gain anatomical information about lung parenchyma and airways is the computed tomography (CT). Segmenting e.g. the tracheo-bronchial tree from these datasets is difficult however. This is due to the small size of the anatomical structures of interest, the oftentimes low contrast between air and lung parenchyma, and the partial volume effect, which decreases this contrast further. Even today, airway segmentation is oftentimes carried out manually, especially in pathological cases with larger malformations. In the last years, a great number of (semi-)automatic methods have been presented in order to facilitate this tedious and time-consuming task (see Section 2). The main intention of the airway segmentation challenge EXACT09 is to give

the possibility to compare these segmentation methods in a meaningful manner (by applying them to the same 20 test datasets).

In this paper, we present our semi-automatic segmentation algorithm and discuss the results achieved within the EXACT09 challenge. The algorithm consists of three basic steps (3D region-growing, 2D wave propagation, and 2D template matching), which are iterated until the desired result is achieved. User interaction is limited to the specification of a seed point inside the easily detectable upper airways and in few cases the decision for a preprocessing step (filtering)[4].

In the following parts of this paper, we briefly review related work in the field of airway segmentation in the next section. Afterwards, we will introduce the hybrid segmentation method, including all its pipeline stages in Section 3. After that, we present and discuss the results achieved within the EXACT09 challenge (Section 4). Finally in Section 5, we present conclusions and point to future research directions.

2 Related Work

Region growing algorithms are a common and frequently-used technique for airway segmentation [1–5]. Here, voxel with gray values within a pre-specified interval and connected to a seed point are included into the segmentation. Besides various advantages (fast, easy to apply), the main drawback is their sensitivity to leakages into the lung parenchyma. Various methods for leakage-avoidance have been presented. Filtering the CT dataset as pre-processing step is a straighforward corrective in this case, although this also implies loss of information on very small bronchi [2]. Gergel et al. introduced adaptive thresholds for leakage prevention with their 3D region growing approach. As soon as a leakage is detected, segmentation is continued with more conservative thresholds avoiding the leakage at that point [1]. Kitasaka et al. controlled leaking and bifurcation problems by a complex use of local volume of interest templates that limit the region growing area [3].

Further techniques used for airway segmentation apply mathematical morphologies [2, 6]. The whole CT dataset is searched for candidate airways with the help of several nonlinear filters. The final segmentation is generated by a reconstruction step that distinguishes correct from false candidates. Despite of the fact that no seed point selection is necessary, a further advantage is that airways, which are not directly connected to a seed point can be detected (useful for pathologic lungs). Main drawback of algorithms using mathematical morphologies is their long runtime.

Combining different methods in an airway segmentation pipeline is another successful approach. Kiraly et al. use a combination of an adaptive 3D region growing, 2D mathematical morphology, and an optional 2D median filter to increase the robustness of the segmentation algorithm while improving the quality of the results [2]. Law and Heng use a combination of region growing and centerline extraction to enhance the understanding of the 3D structure of the bronchial

[4] Other parameters are pre-defined and may be modified by the users.

tree [4]. Tschirren et al. introduced a segmentation based on fuzzy connectivity, i.e. voxels are detected as similar to seed voxels by fuzzy logics. Leakages are prevented by restricting the segmentation to a cylindrical-shaped ROI around bronchi [7]. Graham et al. deploy airway segmentation by detecting shorter bronchi sections, which are represented by surface meshes, and connecting them by interpolation. Strong filtering for the initial segmentation avoids leakages [8]. Finally, another hybrid solution is proposed by Mayer et al. [9], which is the basis for this contribution. It combines region growing with knowledge-based techniques and uses fuzzy logic for the segmentation of the bronchus walls.

3 Airway Segmentation Pipeline

The segmentation pipeline consists of three stages (Fig. 1). In the first stage, the trachea and central bronchi are segmented using standard 3D region growing methods.

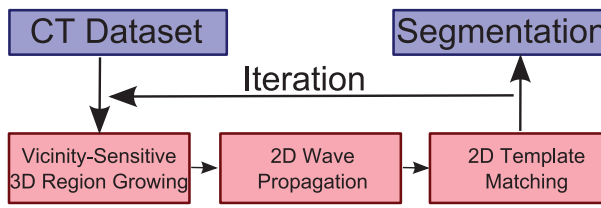


Fig. 1. Segmentation pipeline.

Partial volume effects and limited resolution of the CT scan (which essentially cause this effect) render this method as not satisfactory for segmentations of further generations of the bronchi, since bordering voxels cannot be sufficiently differentiated from tissue voxels. Therefore, a 2D wave propagation is initiated to complete the upper and central branches. Finally, a 2D template matching procedure is used to segment small lumen, which might be only a single voxel large. A feedback loop of the whole pipeline repeats the stages until no meaningful additions can be made to the previous segmentation (Fig. 1). Figure 5 shows the final results of five iterations. However, some datasets might require up to 15 iterations.

3.1 Vicinity-Sensitive 3D Region Growing

The intensity values of CT dataset of the thorax can be divided into three categories (Fig. 2). Values below -950 Hounsfield units (HU) can be classified as **definitely airway** and values above -775 HU as **non-airway**. Voxels with values in between (in the isovalue interval from -950 HU to -775 HU) can belong

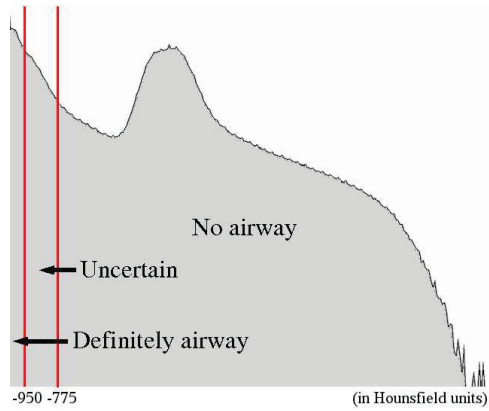


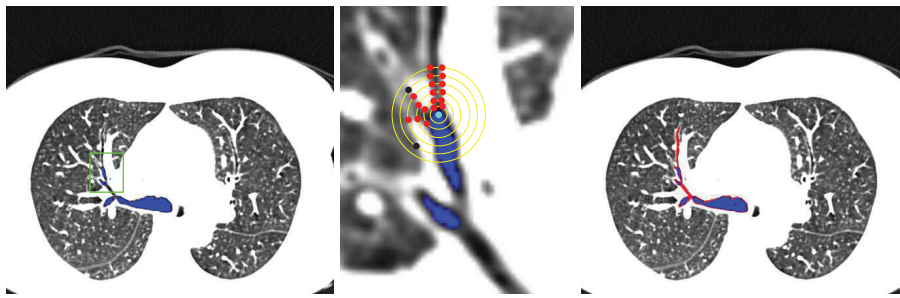
Fig. 2. Logarithmic histogram of CT thorax dataset.

to airway or tissue. They are classified as **uncertain** and need to be investigated further.

Based on this analysis, the 3D region growing algorithm extracts all voxels which are definitely airway, starting at the user-defined seed point in the trachea. To prevent the leaking into the parenchyma of the lungs in smaller airways (i.e., in emphysema), we use a masking technique from texture analysis; if the average gray value of a $3 \times 3 \times 3$ voxel cube centered at the current voxel is within the save range (below -950 HU), we consider this voxel as being part of the airway. Otherwise, the respective voxel is not classified as airway in this stage. While this masking technique prevents leakage, it also impedes the segmentation of smaller bronchi. However, we usually accomplish the segmentation of the bronchial tree up to the fifth generation, whereas the bordering voxels are often not included, since their voxel values belong to the uncertain voxel value interval (see Fig. 3 and Fig. 5a/b). In the second iteration of the segmentation pipeline, the 3D region growing algorithm runs with the same threshold on the bordering voxels of the previously selected voxels.

3.2 2D Wave Propagation

Starting from segmented voxels of the previous step, 2D wave propagation tries to reconstruct bronchi walls within a single CT slice. It starts at each boundary voxel of the airway voxels from 3D region growing and propagates waves to detect the walls of the bronchi (Fig. 3b). Voxels at position X in the **uncertain** areas are classified by fuzzy logic rules that consider the density value V(X) (in Hounsfield units) , the largest local N4 neighborhood (in 2D) gradient G(X), and if voxels in the local N4 neighborhood are already classified as wall pixels



(a) (b) (c)

Fig. 3. Completion of bronchi walls; (a) shows the result of the 3D region growing (blue). The green rectangle marks the zoomed area shown in (b). (b) shows the wave propagation in progress. The cyan boundary voxel is chosen as starting point. The yellow circles mark the propagated waves and the red points mark the airway candidate voxels. The black points failed the leaking test, since the number of voxels of that wave was increasing too fast. Note that the actual waves have a Rhombus-like shape, driven by the N4-neighborhood. (c) shows the completion of that segmentation by 2D wave propagation (red). The voxels (within the body) in the iso-range of **definitely airway** are marked in blue, of the **uncertain** range in black/grey, and of the **no airway** range in white.

(no airway) in a previous wave W(X):

$$f_{wave}(X) = c_v * V(X) + c_g * G(X) + c_w * W(X), \quad (1)$$
$$with\ c_v = 1, c_g = 1, c_w = 0.75.$$

where V(X) and G(X) are mapped into the closed interval [1.0, 0.0], and W(X) is either 1 – if there is a classified wall pixel in the N4 neighborhood — or 0 – otherwise. Essentially, if $f_{wave}(X) \geq c_{wall}$, the voxel is classified as wall[5].

Critical to the wave propagation is the evaluation of the classified airway areas, if they really belong to the airways. To achieve this goal, the additional voxels segmented by each wave are monitored by a protocol that verifies the shape and size of each bronchus candidate, using a set of default parameters (Fig. 3b). As metric, we count the number of voxels selected by the n waves propagating within a plane (BPS_n for BronchusPlaneSize), and the wave diameter (WD_n) of the current wave n as the number of selected voxels of wave n. Furthermore, we define the average number of voxels of the first n waves (AWD_n for Average Wave Diameter).

Segments of the tracheo-bronchial tree are identified by sequences of as airway classified voxels in a wave. Figure 3b shows two sequences marked by the red points, thus depicting a bifurcation. The shape rules essentially assume that no wave detects segment splits in more than two subsequent segments at a bifurcation. A third segment (of not yet selected voxels) in one 2D wave propagation

[5] Typically, $c_{wall} = 1.74$.

test (in one slice) is henceforth considered as leakage into the lungs and is considered invalid. At each bifurcation, the segment identification process starts again recursively.

As closer examination of previously examined CT thorax datasets showed, two very close bifurcations were never located close enough to be detected as a third segment by the 2D wave propagation, thus they were not falsely identified as leakage.

$$WD_n > d_{max} \tag{2}$$

$$WD_n/WD_{n-1} > d_{WD_{ratio}} \tag{3}$$

$$BPS_n > d_{size} \tag{4}$$

$$AWD_n - AWD_{n-1} > d_{AWD_{current}} \tag{5}$$

$$AWD_{max} - AWD_{min} > d_{AWD_{longterm}} \tag{6}$$

The size rules limit the growing of the wave propagation[6]. If the diameter of a bronchi candidate exceeds a certain size (Equation 2), or if the wave diameter is increasing too fast from the previous wave (Equation 3, see also black points in Fig. 3b), the respective segment recursion is terminated and the results are considered as leakages (invalid). Furthermore, if candidates grow spontaneously (while shrinking before) or the overall in plane voxel size BPS_n of the candidate becomes unrealistically large (Equation 4), the recursion is again terminated and the results are set to leakages (invalid). The last two rules (Eqn. 5 and 6) test the current and long-term growth of the wave front. Specifically, they test if the segments are shrinking (as assumed) or growing. The protocol starts testing after the first three waves, since they frequently show an unstable behavior. During the wave propagation, all invalid results are removed from the segmentation. However, initial correct results (i.e., for the first p waves) are preserved.

To follow a bronchus through several slices, virtual waves are propagated in neighboring slices. If one of these virtual waves is similar to the shape and size of the wave propagation in the current slice, another recursive wave propagation in the neighboring slice is initiated. Specifically, the recursive testing of waves in neighboring slice is initiated only for no-branching segments that have classified wall elements from 2D wave propagation. Furthermore, these wall elements may only differ by one voxel to the wall elements of the new neighboring slice segment.

Similar to the first step of the pipeline, 2D wave propagation uses almost the same parameters in the subsequent iteration; only the peripheral bronchi diameter is reduced since the lower airways (higher generations) only grow smaller.

3.3 2D Template Matching

Without the used careful validity testing, the previous two stages would leak into the surrounding area, if the airways become too small to be picked up,

[6] We use $d_{max} = 6.1mm,\ d_{WD_{ratio}} = 1.75,\ d_{size} = 500mm^2,$ $d_{AWD_{current}} = 1.13,\ d_{AWD_{longterm}} = 3.0.$

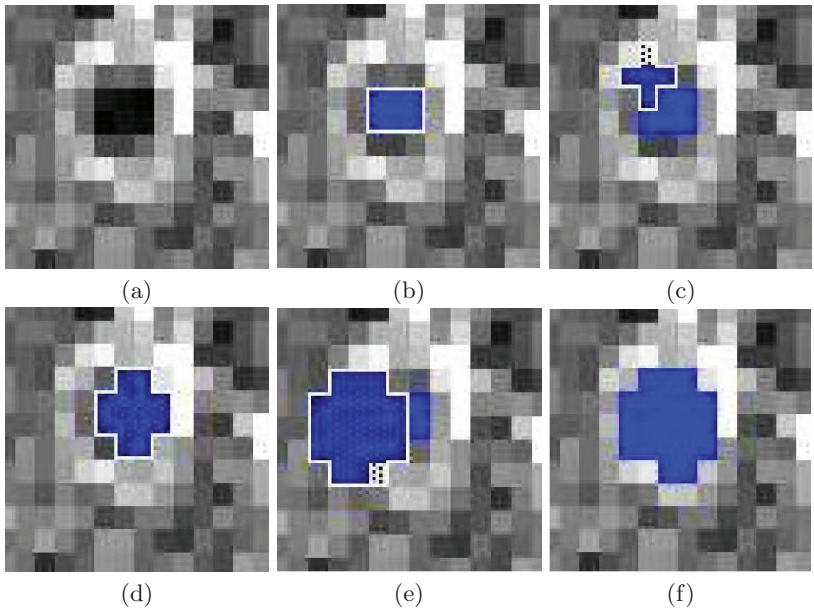


Fig. 4. Template matching: (a) shows a peripheral airway. 2D seeding is started on pixels of the category **uncertain**. The result is marked blue (b). From there the seeding area templates (white polygon with black dot pattern) are formed and tested on different locations around the seeding. The seeding is repeated on the category uncertain and template (marked by a pattern, (c) to (e)). (f) shows the best result after the classification.

in particular in areas where the airways might have the size of only one voxel. To select these voxels, but still prevent the leaking, we apply a 2D template matching technique that evaluates the candidate area below templates with the isovalue category **uncertain** (between -950 HU and -775 HU). This stage is organized in two steps; the first step establishes templates that are used in the second step to evaluate the local voxel neighborhood.

First, 2D template matching applies 2D region growing starting from the boundary voxels of the previous segmentations (Fig. 4). The thresholds are varied – from the upper threshold of the **uncertain** isovalue interval (-775 HU) – until the number of selected voxels is below the critical limit (i.e., 35 voxels), since it can be assumed that they did not leak out. Based on this selected voxel area, circular templates of varying sizes are generated.

In the second step, we apply a 2D region growing and use the templates to differentiate the thresholds; below the template, we are using the upper **uncertain** threshold (-775 HU), while we are using the original template threshold outside of the template. By moving the templates over the local area, we generate various segmentation candidates (see Fig. 4b-e) which are again evaluated by a set of fuzzy rules. This time, we consider the average density value $\bar{V}(X)$ of

the template area and the average (gradient range is clamped in order to reduce data artifacts) gradient $G(X)$ to the surrounding voxels in the N8 neighborhood (within a single slice). The best possible result is then selected and added to the segmentation (Fig. 4 and Fig. 5d).

$$f_{template}(X) = c_{ave} * \bar{V}(X) + c_{grad} * G(X), \quad with\ c_{ave} = 0.25\ and\ c_{grad} = 0.75. \tag{7}$$

Here, $\bar{V}(X)$ is mapped from [-1000,-775] to [0.0,1.0], and $G(X)$ is mapped into the range [0.0, 1.0]. Illustratively, this means that accepted candidates have a low average density value and a high boundary contrast. The candidate with the largest $f_{template} \geq 0.7$ (Eqn. 7) is accepted as airway (Fig. 4f and Fig. 5d). However, if the size of the template controlled area is larger than twice as much as for the previous slice, a leak-out is assumed, thus the area of the current slice is assumed invalid.

In the subsequent iterations, voxels which have already been unsuccessfully tested for inclusion, are excluded from template matching. This is mainly to save time – 2D template matching is the single most time consuming stage of the segmentation pipeline – and they usually do not contribute in later iterations.

4 Application to EXACT09 datasets

When applying our semi-automatic segmentation method, user interaction is required for setting the seed point in the trachea. All parameters (thresholds, propagation diameter, etc.) are pre-specified. These default parameters (see Section 3), which were fixed based on previously segmented CT datasets and validated with the provided training data, also worked very well with the majority of the test datasets. An optimization of the parameters for individual datasets is possible, but improves the segmentation results only slightly compared to the standard setting. We consider this as stability feature of our approach.

In several cases, we faced the problem that due to noise effects (especially in low-dose CT datasets), the initial vicinity-sensitive 3D region-growing showed no satisfying results. A slight adjustment of the segmentation thresholds might solve this problem (as e.g. in case 26). However, these thresholds are based on expert knowledge, so increasing these parameters too much easily leads to false positives, when uncertain or tissue voxels are classified as airway. Thus, in these cases a preprocessing step, i.e. the application of a gentle Gaussian filter (kernel size of 3 voxels), was included into the pipeline. Although filtering also implicates a loss of information concerning the lower and smaller bronchi, it turns out to be the better choice than adjusting thresholds, risking leakages, and thereby a higher false positives rate. Altogether, this pre-processing step allowed fairly good segmentations compared to the problems that arose before.

The segmentation results were evaluated according to two main categories: the overall segmentation result and the leakage robustness (see Table 1). Concerning the segmentation sensitivity, our method detected 41.7% of the branches

Table 1. Evaluation measures for the twenty cases in the test set.

	Branch count	Branch detected (%)	Tree length (cm)	Tree length detected (%)	Leakage count	Leakage volume (mm^3)	False positive rate (%)
CASE21	64	32.2	35.5	32.1	0	0.0	0.00
CASE22	144	37.2	98.9	29.9	0	0.0	0.00
CASE23	158	55.6	112.3	43.1	5	89.8	0.70
CASE24	112	60.2	83.4	51.3	2	8.2	0.04
CASE25	151	64.5	110.5	43.8	4	50.7	0.21
CASE26	54	67.5	38.4	58.4	3	297.1	4.85
CASE27	52	51.5	35.7	44.0	0	0.0	0.00
CASE28	89	72.4	63.2	57.6	0	0.0	0.00
CASE29	58	31.5	36.1	26.2	0	0.0	0.00
CASE30	98	50.3	67.6	44.2	1	1.8	0.03
CASE31	61	28.5	38.6	22.0	1	6.7	0.08
CASE32	64	27.5	46.8	21.5	2	108.6	0.93
CASE33	55	32.7	38.4	26.1	0	0.0	0.00
CASE34	139	30.3	84.8	23.7	4	37.6	0.20
CASE35	180	52.3	117.2	37.9	13	155.4	0.93
CASE36	69	19.0	60.8	14.7	0	0.0	0.00
CASE37	53	28.6	43.0	24.2	0	0.0	0.00
CASE38	40	40.8	30.6	46.1	0	0.0	0.00
CASE39	101	19.4	78.9	19.3	0	0.0	0.00
CASE40	127	32.6	93.4	24.1	3	28.6	0.20
Mean	93.5	41.7	65.7	34.5	1.9	39.2	0.41
Std. dev.	43.0	16.2	29.6	13.2	3.1	74.8	1.09
Min	40	19.0	30.6	14.7	0	0.0	0.00
1st quartile	55	28.6	38.4	23.7	0	0.0	0.00
Median	79	35.0	62.0	31.0	1	0.9	0.01
3rd quartile	144	60.2	98.9	46.1	4	89.8	0.70
Max	180	72.4	117.2	58.4	13	297.1	4.85

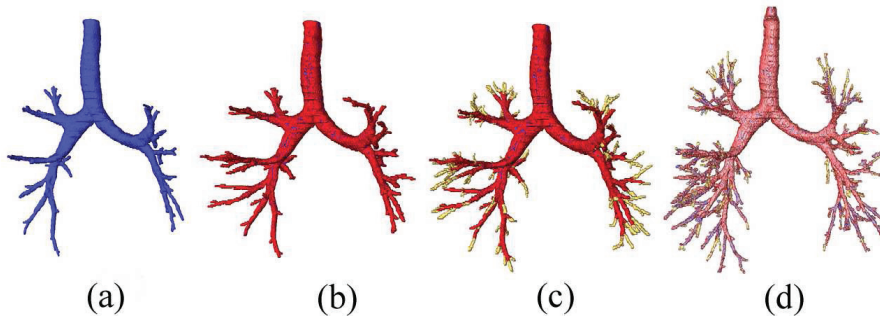


Fig. 5. Three segmentation stages: (a) region growing, (b) wave propagation, (c) template matching, (d) final segmentation result

and 34.5% of the tree length (length of the centerlines of detected branches). Both compared to the gold standard, which is the airway segmentation result carried out manually by radiologists and other medical experts. With that, our methods belongs to the lower midrange, when compared to all other 14 algorithms, and ranks second in the group of semi-automatic methods. A problem reducing the sensitivity of our method, were larger gray values occuring in the bronchi due to pathologies (e.g. caused by secretion). These voxel are then classified as lung parenchyma and interrupt the segmentation (see Figure 6).

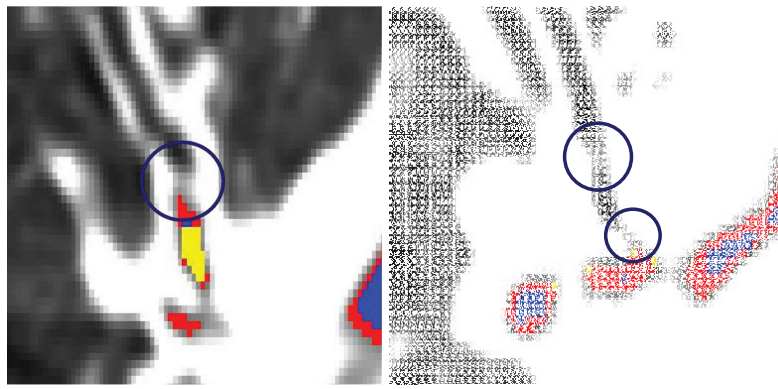
The strength of our algorithm, however, is leakage-control. Only 0.41% of the detected branches were false positives (in Table 1 referred to as *false positive rate*) and with that result we achieved rank three in total. The positive predictive value, which measures the accuracy that existing branches are detected (correctly found bronchi branches / (correctly found bronchi branches + found false positive bronchi branches)), is 98.0% and in 50% of our segmentations, no leakage occured at all.

In the segmentation of airways in CT datasets, a satisfying trade-off between detecting a maximal number of bronchi and avoiding leakages has to be found. The evaluation within the EXACT09 challenge shows, that our segmentation method belongs to the more conservative techniques, where leakage-avoidance has a very high priority.

As noted before, usually five to seven iterations are sufficient for a segmentation of the tracheo-bronchial tree. Overall, this corresponds to a typical segmentation duration between 10 and 30 seconds on a PC with a Intel Core 2 Duo processor, each core running at 2.4 GHz.

5 Conclusions and Future Work

In this paper, we presented a semi-automatic segmentation method with a pipeline of three main steps, which is iteratively applied. These main steps are 3D region-



(a) (b)

Fig. 6. Areas with higher gray values (marked by circles) inside the bronchi interrupt the segmentation. Examples from Case36 (a) and Case39 (b).

growing, 2D wave propagation, and 2D template matching, all of which are implemented with special focus on leakage-prevention (see Section 3). The results within the EXACT09 airway segmentation challenge justify this effort with a very low false positive rate and a small leakage volume. The drawback of the powerful leakage-control is, however, the reduced sensitivity concerning the overall segmentation performance.

A straightforward advancement of our method is stepping from 2D to 3D for the wave propagation and the template matching. At present, these two pipeline stages work on individual slice images. Thus, detecting bronchi located orthogonal to these slices is usually not possible within the same iteration. It turned out, that the subsequent vicinity-sensitive 3D region growing – followed by the other 2D pipeline stages – is able to compensate for this effect in many cases. Hence, implementing 3D wave propagation and 3D template matching will primarily show an improvement concerning runtime (by achieving the same result in fewer iterations), but beyond that we also except some better identification of smaller bronchi.

References

1. Gergel, I., Wegner, I., Tetzlaff, R., Meinzer, H.P.: Zweistufige Segmentierung des Tracheobronchialbaums mittels iterativen adaptiven Bereichswachstumsverfahren. In: Proc. of Workshop Bildverarbeitung für die Medizin. (2009) 56–60
2. Kiraly, A., Higgins, W., McLennan, G., Hoffman, E., Reinhardt, J.: Three-Dimensional Human Airway Segmentation Methods for Clinical Virtual Bronchoscopy. Academic Radiology **9**(10) (2002) 1153–1168

3. Kitasaka, T., Mori, K., Hasegawa, J., Toriwaki, J.: A Method for Extraction of Bronchus Regions from 3D Chest X-ray CT Images by Analyzing Structural Features of the Bronchus. Forma **17**(4) (2002) 321–338
4. Law, T., Heng, P.: Automatic Centerline Extraction for 3D Virtual Bronchoscopy. In: Proc. of MICCAI. Volume 1935. (2000) 786–795
5. Mori, K., Hasegawa, J., Suenaga, Y., Toriwaki, J., Anno, H., Katada, K.: Automated Labeling of Bronchial Branches in Virtual Bronchoscopy System. In: Proc. of MICCAI. Volume 1496. (1998) 870–878
6. Fetita, C., Preteux, F., Beigelman-Aubry, C., Grenier, P.: Pulmonary Airways: 3-D Reconstruction form Multislice CT and Clinical Investigation. In: IEEE Trans. on Medical Imaging. Volume 23. (2004) 1353–1364
7. Tschirren, J., Hofmann, E., McLennan, G., Sonka, M.: Intrathoracic Airway Trees: Segmentation and Airway Morphology Analysis From Low-Dose CT Scans. In: IEEE Trans. on Medical Imaging. (2005) 1529–1539
8. Graham, M., Gibbs, J., Higgins, W.: Robust System for Human Airway-Tree Segmentation. In: Proc. of SPIE Medical Imaging. Volume 6914. (2008) 69141J
9. Mayer, D., Bartz, D., Fischer, J., Ley, S., del Rio, A., Thust, S., Kauczor, H.U., Heussel, C.P.: Hybrid Segmentation and Virtual Bronchoscopy Based on CT Images. Acad Radiol **11** (2004) 551–565

Results of Applying Two-Pass Region Growing Algorithm for Airway Tree Segmentation to MDCT Chest Scans from EXACT Database

Anna Fabijańska

Department of Computer Engineering, Technical University of Lodz,
18/22 Stefanowskiego Str., 90-924 Lodz, Poland

an_fab@kis.p.lodz.pl

http://www.kis.p.lodz.pl

Abstract. This paper addresses problem of CT based investigation of pulmonary airways. Especially algorithm for airway tree segmentation is introduced. The algorithm is based on 3D seeded region growing. In the proposed method region growing is repeated twice with different settings. First, in order to avoid leakages into the lungs, algorithm is performed with very strict criteria. Next it works only in the places of interest which are found by morphological gradient and threshold. Results of applying the proposed method to several CT chest scans selected randomly from EXACT database are presented and discussed. Moreover comparision with the ground truths, provided by EXACT organisers, is given by means of the evaluation measures.

Key words: Airway Tree Segmentation, Bronchial Tree, Clinical Investigation, CT bronchography, Pulmonary Imaging, Region Growing, Multidetector Computed Tomography, MDCT, X-Ray CT

1 Introduction

Multidetector computer tomography is a powerful technique which produces detailed three-dimensional images of the inside of a human body. Due to its effectiveness computer tomography (CT) is successfully used in many clinical applications and physiological studies to investigate changes due to various diseases.

One of the examples of computer tomography usage is an investigation of pulmonary disorders. CT chest scans are especially helpful in diagnosis of chronic obstructive pulmonary disease (COPD) which is a common name for pathological changes due to different combinations of airway diseases and asthma which are characterized by airflow limitation [1][2][3]. Quantitative description of an airway wall, especially the thickness of airway walls and diameter of an airway lumen, provide important information about pulmonary diseases. Therefore reliable segmentation of an airway tree from volumetric computer CT data sets

is the most important step in clinical applications for measuring and characterizing airway data [4]. Still segmentation of an airway tree from CT chest scans is very challenging. Problems are caused by inhomogenity of a bronchial lumen, adjacency of the blood vessels, and changes of intensities along airway walls. Moreover, in case of airway tree segmentation region growing is prone to leakages into lungs.

Different algorithms for airway tree segmentation from CT chest scans have been proposed in the literature. Mostly they act on 3D volumetric datasets and are region growing based approaches utilizing different techniques in order to diminish leakage probability. The most popular algorithms can be classified into one of the following groups:

- **Rule-based techniques** which utilize anatomical knowledge about airways and blood vessels [5] [6] [7].
- **Morphological techniques** using grayscale morphological reconstruction to identify local extremes in the image [8] [9] [10].
- **Wave front propagation techniques** which propagate waves to detect walls of the bronchi [8] [11] [12].
- **Template matching techniques** that search consecutive slices for oval dark rings surrounding brighter areas (airways) or dark solid oval areas (adjacent blood vessels) [13] [14].
- **Fuzzy techniques** which utilize fuzzy rules to avoid leakages and diminish number of falsely detected bronchi [15] [16].

In this paper a new approach for airway tree segmentation is presented. The method uses modified 3D region growing algorithm where the growth of an airway tree is guided and constrained by morphological gradient. In consequence leaks into the lungs are avoided.

The outline of this paper is as follows. In Section 2 a description of datasets used in this work is given. Description of the authors' two-pass region growing algorithm (TPRGA) for airway tree segmentation is given in detail in Section 3. In Section 4 results of the proposed algorithm are demonstrated using datasets from EXACT (Extraction of Airways from CT) [20] database. Finally obtained results are analysed discussed in Section 5. Section 6 concludes the paper.

2 Input Data

3D volumetric CT chest scans of forty patients provided by EXACT organizers were examined. They were acquired at different sites using several different scanners, scanning protocols, and reconstruction parameters. The dataset ranged from clinical dose to ultra low dose scans, from healthy volunteers to patients with severe lung disease, and from full inspiration to full expiration. Two datasets were provided: the training set (CASES 1-20) and the testing set (CASES 21-40).

The slices were provided with 16-bit resolution and stored as signed 16-bit monochromatic images of the resolution 512x512 pixels. Individual slices were

stacked into a 3D space representing volumetric data set. Exemplary CT slices (displayed using the same window and level settings) are presented in Figure 1.

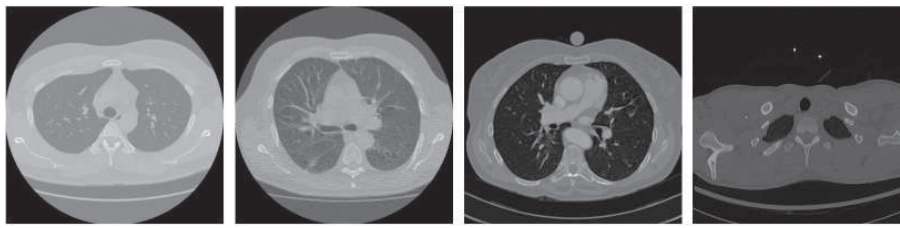


Fig. 1. Exemplary CT slices from examined datasets.

Single CT chest scan consists of about 450-600 planar transverse slices. Consecutive slices combined into a stack compose volumetric data set in three dimensional (3D) space.

3 Problems with Region Growing

An airway tree, which conducts air into the lungs, is build from an airway lumen surrounded by high density vascular wall. Both an airway lumen and lungs are filled with air.

Due to differences in their densities on CT slices an airway wall appears significantly brighter than the area of airways and lungs. Therefore in the ideal case using 3D seeded region growing with a seed point located at the first slice in the centre of the trachea it should be possible to segment complete airway tree from a volumetric CT data set. This idea is presented in Figure 2.

In practice airway walls often appear broken. In consequence region growing leaks from an airway lumen into the lungs which are of similar intensity. Region growing algorithm leaking into the lungs is presented in Figure 3.

4 The Proposed Approach

The proposed airway tree segmentation algorithm, as most approaches to the considered problem, is based on seeded 3D region growing. However during airway lumen segmentation region growing is applied twice. Moreover, leak prevention mechanism is applied in order to avoid the algorithm to consider voxels that are part of the lung parenchyma. Successive steps of the algorithm are as follows:

1. Contrast Enhancement.
2. Detection of the Seed Point.
3. First Pass of Region Growing.

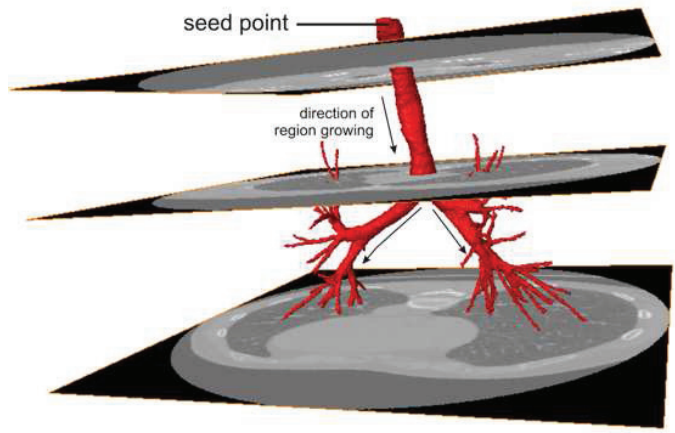


Fig. 2. The idea of an airway tree segmentaion by 3D region growing.

4. Computation of Morphological Gradient.
5. Thresholding of Morphological Gradient.
6. Second Pass of Region Growing.

Step 1: *Contrast Enhancement.* Before region growing is applied contrast of the input dataset is improved. In this stage histogram of the whole volume is normalized [17]. After normalization there are 0.5% of saturated voxels in the CT dataset.

Step 2: *Detection of the Seed Point.* A seed point for the region growing is determined automatically. It is defined as a voxel located inside the trachea.

Trachea is easy to find. Its cross sections are imaged as large oval areas and are present in the central part of the first slices of the CT chest scan. They are also significantly darker than their surrounding. Therefore in case of the presented method location of the trachea is defined using template matching method [17] searching for a dark oval area. The search starts from the first slice and is continued in consecutive ones until the oval of tracheal tube is found. Seed point is defined as its center of gravity [17].

Step 3: *First Pass of Region Growing.* When seed point is determined first pass of 3D region growing is performed on the normalized CT dataset. In order to avoid leakages into the lungs very strict criteria are used in this stage while joining consecutive voxels. The algorithm classifies the current voxel as voxel belonging to the airway lumen if its intensity and intensity of its all closest (connected) neighbours differ from the average intensity of voxels already classified to airway lumen not more than T%. Value of T is determined automatically on the first run. It equals to the lowest value for which region growing starts up. The value of T is then decreased by half after trachea is segmented. This allows

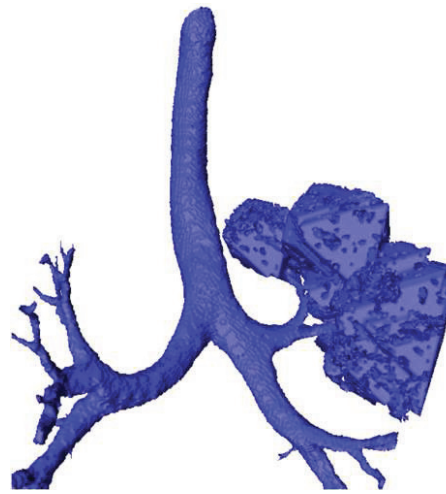


Fig. 3. Region growing leaking into the lungs.

to avoid leakages into the lungs as fluctuations of voxel intensities in case of distal bronchi are smaller than in the area of the trachea.

Step 4: *Computation of Morphological Gradient.* The following step of the algorithm aims at highlighting information connected with airway walls. In order to do so 3D morphological gradient [17] is calculated based on original (i.e. not normalized) data set.

Morphological gradient (which is a difference between results of grayscale dilatation and grayscale erosion) emphasizes sharp gray level transitions, which are also connected with airway walls. Those areas of the highest gradient which are connected with airway tree built in the third step of the algorithm define possible locations of the distal bronchi.

It is important to perform this step on the unnormalized CT dataset. Otherwise noise highlighted by histogram stretching can seriously hamper accurate segmentation of complete airway tree.

Step 5: *Thresholding of Morphological Gradient.* In order to determine areas of the highest gradient its image is thresholded with global threshold. For threshold selection ISODATA method [18] [19] is applied. Airways are supposed to be located in those binary areas which are connected with airway tree built in the second step of the algorithm.

Step 6: *Second Pass of Region Growing.* The second pass of region growing is performed again on normalized data set but not everywhere. Places of interest are determined by morphological gradient.

Region growing starts from previously segmented airway tree. It is guided and constrained by those areas of the highest gradient (determined by thresholding

in previous step) which are connected to airway tree built in the third step of the algorithm. Successive voxels are joined to airway tree if both of the following constraints are fulfilled:

- current voxel is situated in the area of the highest gradient determined in the previous step;
- intensity of the current voxel differs from the average intensity of voxels classified to airway lumen not more than $2T\%$. Where value T is remembered from the third step of the algorithm (the lowest value for which region growing starts up) and changes during algorithm performance in the way as during the first pass of region growing.

Note: All algorithm parameters were determined only using the training data. Value of T selected as described above in case of all tested 40 datasets allowed to avoid the leakages into lungs. However, in some cases it was possible to tune the algorithm performance and obtain better accuracy of the results by increasing value of T manually. Therefore the presented algorithm was classified as *semi-automatic*.

5 Results and Discussion

Results obtained by the proposed airway tree segmentation algorithm are presented in Table 1 and in Figure 4.

Table 1 presents outcomes of evaluation of airway tree segmentations obtained by the proposed algorithm from the test datased. The evaluation was performed manually by EXACT organisers. Obtained segmentations were compared with the ground truth build in the way as explained on EXACT website [20] in *Information* section.

The following evaluation measures were used for results evaluation:

- **Branch count** - the number of branches that were detected correctly (i.e. branches with centerlines longer than 1 mm).
- **Branch detected** - the fraction of branches that were detected, with respect to the branches present in the ground truth.
- **Tree length** - the sum of the length of the centerlines of all correctly detected branches.
- **Tree length detected** - the fraction of tree length in the ground truth that was detected correctly.
- **Leakage count** - the number of unconnected groups of "correct" regions that are neighboring with a "wrong" region.
- **Leakage volume** - the volume of regions that are wrongly detected.
- **False positive rate** - the fraction of the volume of regions that are detected wrongly over the volume of all detected regions.

Trachea was excluded from the branch length and branch count related measurements. For the lekage based measures, both trachea and main bronchi were excluded.

Table 1. Evaluation measures for the twenty cases in the test set.

	Branch count	Branch detected (%)	Tree length (cm)	Tree length detected (%)	Leakage count	Leakage volume (mm^3)	False positive rate (%)
CASE21	115	57.8	61.3	55.5	6	307.2	2.61
CASE22	84	21.7	59.4	18.0	13	510.9	2.82
CASE23	84	29.6	59.9	23.0	1	0.4	<0.01
CASE24	53	28.5	44.2	27.1	0	0.0	0.00
CASE25	61	26.1	48.1	19.1	0	0.0	0.00
CASE26	44	55.0	29.1	44.3	3	196.2	2.68
CASE27	26	25.7	20.4	25.2	0	0.0	0.00
CASE28	80	65.0	56.4	51.5	0	0.0	0.00
CASE29	104	56.5	66.7	48.3	1	1.4	0.01
CASE30	118	60.5	82.9	54.3	3	84.2	0.56
CASE31	92	43.0	60.7	34.6	3	277.1	1.53
CASE32	65	27.9	45.3	20.8	0	0.0	0.00
CASE33	77	45.8	51.8	35.2	5	193.7	1.96
CASE34	173	37.8	119.0	33.3	1	6.3	0.02
CASE35	43	12.5	30.8	10.0	0	0.0	0.00
CASE36	33	9.1	24.9	6.0	0	0.0	0.00
CASE37	74	40.0	54.7	30.8	0	0.0	0.00
CASE38	49	50.0	36.3	54.6	1	0.4	<0.01
CASE39	60	11.5	45.0	11.0	5	622.0	5.77
CASE40	116	29.8	90.6	23.4	3	125.6	0.48
Mean	77.5	36.7	54.4	31.3	2.3	116.3	0.92
Std. dev.	35.4	17.1	23.5	15.7	3.2	184.6	1.54
Min	26	9.1	20.4	6.0	0	0.0	0.00
1st quartile	49	25.7	36.3	19.1	0	0.0	0.00
Median	76	33.8	53.3	29.0	1	0.9	0.01
3rd quartile	115	56.5	66.7	51.5	5	277.1	2.61
Max	173	65.0	119.0	55.5	13	622.0	5.77

In Figure 4 3D views for twelve CT datasets selected randomly from EXACT database are shown. Cases from both train and test datasets are considered in the figure. Blue colour is assigned to airway tree extracted by the first pass of the region growing. Airways segmented during the second pass of the algorithm are marked with red colour. CASE ID is indicated under each subimage.

Presented results prove that the proposed airway tree segmentation method succeeded in all 40 cases. An airway tree was segmented without significant leaks into the lungs.

The outcomes of results evaluation show that the method extract airway trees with the average accuracy of branch detection equal to 37% and average fraction of tree length equal to 36%. The average rate of falsely detected branches is less than 1% with the average leakage count equal to 2.3.

The accuracy of the proposed airway tree segmentation algorithm can be considered satisfying but also convinces to further improvements of the method.

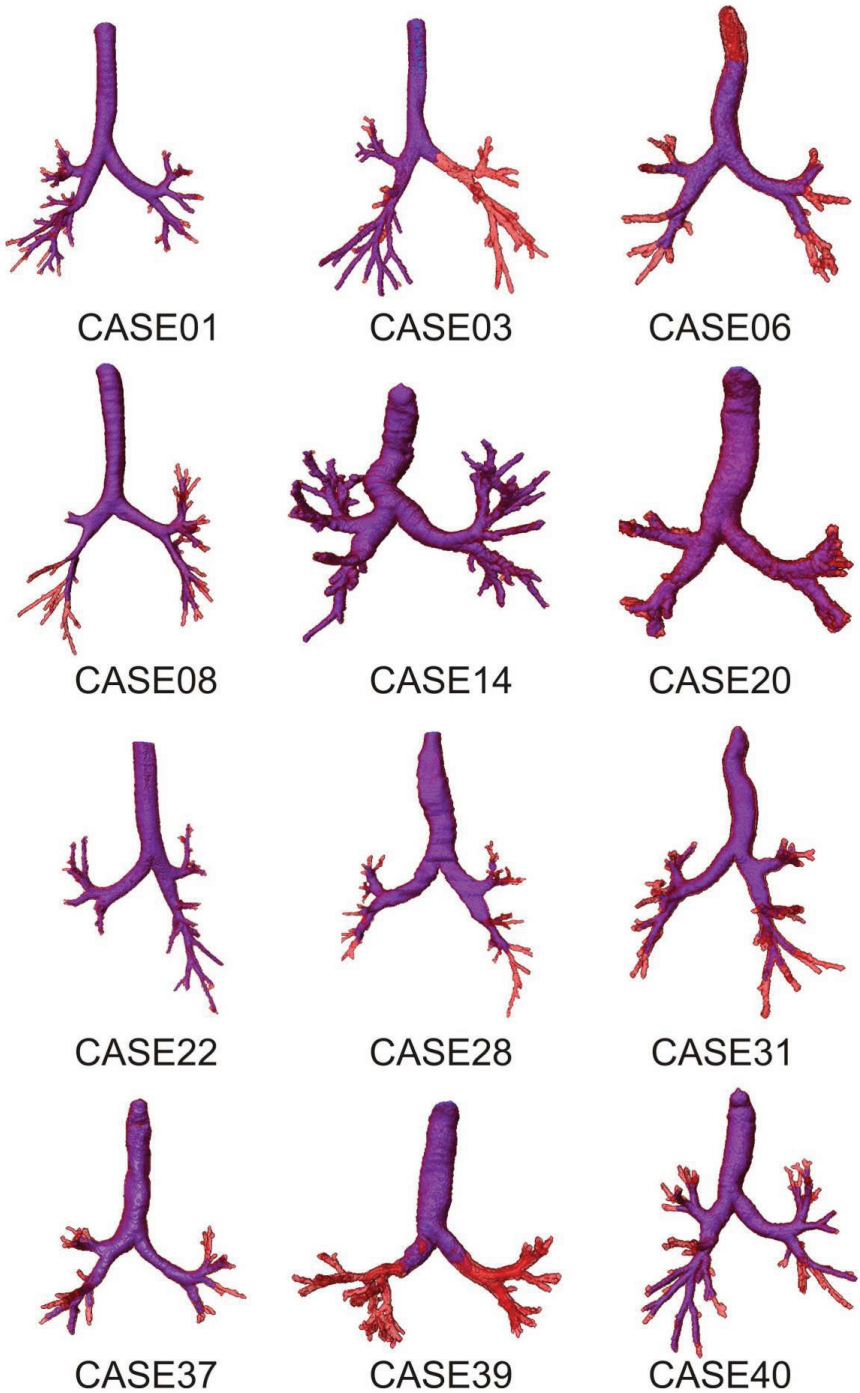


Fig. 4. Results of airway tree segmentation using the proposed method. Blue colour is assigned to airway tree extracted by the first pass of the region growing. Airways segmented during the second pass of the algorithm are marked with red colour. CASE ID is indicated under each image.

The runtime of the proposed algorithm for data set consisting of about 500 16-bit slices of the resolution 512x512 pixels measured on 1.66 GHz Intel single CPU system is not more than 10 minutes. However the method is fully parallelizable and can be adapted to multiprocessor systems (for example using OpenMP API [21]) which are supposed to cut down the runtime of the algorithm.

6 Conclusions

In this paper the problem of an airway tree segmentation from CT chest scans was considered. Especially the new method for airway tree segmentation was introduced. The algorithm is a 3D approach which extracts airway trees from volumetric CT chest scans using 3D region growing guided and constrained by a morphological gradient. This allows to diminish probability of leakages into the lungs.

The algorithm was tested on forty CT chest scans from EXACT database. They were obtained using different protocols, from patients with different level of pathological changes due to lung diseases. Results of applying the method to selected datasets were presented and discussed.

In case of all tested datasets proposed algorithm succeeded. The accuracy of obtained results (i.e. number of extracted generations of bronchi) depends on level of pathological changes in lung area. However in case of majority of datasets up to 6 generations of bronchi was extracted by the algorithm. This accuracy is sufficient for clinical applications [22] [23].

7 Acknowledgements

Anna Fabijańska is a scholarship holder of project entitled "Innovative education ..." supported by European Social Fund.

References

1. American Thoracic Society Statement: Standards for the diagnosis and care of patients with chronic obstructive pulmonary disease. Am. J. Respir. Crit. Care Med. 152, S77–121 (1995)
2. Nakano, Y., Muro, S., Sakai, H., Hirai, T., Chin, K., Tsukino, M., Nishimura, K., Itoh, H., Par, P.D., Hogg, J.C., Mishima, M.: Computed tomographic measurements of airway dimensions and emphysema in smokers correlation with lung function. Am. J. Respir. Crit. Care Med. 162(3), 1102–1108 (2000)
3. Berger, P., Perot, V., Desbarats, P., Tunon-de-Lara, J.M., Marthan, R., Laurent, F.: Airway wall thickness in cigarette smokers: quantitative thin-section CT assessment. Radiology 235(3), 1055–1064 (2005)
4. Reilly, J.: Using computed tomographic scanning to advance understanding of chronic obstructive pulmonary disease. Proc. Am. Thorac. Soc. 3(5), 450–455 (2006)
5. Sonka, M., Park, W., Hoffman E.: Rule-based detection of intrathoracic airway trees. IEEE Tran. Med. Imag. 15(3), 314–326 (1996)

6. Zrimec, T., Busayarat, S.: A system for computer aided detection of diseases patterns in high resolution CT images of the lungs. In 20th IEEE Int. Symp. Computer-Based Medical Systems, pp. 41–46, IEEE Press, New York (2007)
7. Brown, M.S., McNitt, M.F., Mankovich, N.J., Goldin, J., Aberle, D.R.: Knowledge-based automated technique for measuring total lung volume from CT. Proc. SPIE 2709, 63–74 (1996)
8. Felita, C.I., Preteux, F., Beigelman-Aubry, C., Grenier, P.: Pulmonary airways: 3-D reconstruction from multislice CT and clinical investigation. IEEE Trans. Med. Imag. 23(11), 1353–1364 (2004)
9. Aykac, D., Hoffman, E.A., McLennan, G., Reinhardt, J.M.: Segmentation and analysis of the human airway tree from three-dimensional X-ray CT images. IEEE Trans. Med. Imag. 22(8), 940–950 (2003)
10. Pisupati, C., Wolf, L., Mitzner, W., Zerhouni, E.: Segmentation of 3D pulmonary trees using mathematical morphology. In Mathematical morphology and its applications to image and signal processing. Dordrecht, The Netherlands: Kluwer Academic Publishers, pp. 409–416 (1996)
11. Wood, S., Zerhouni, A., Hoffman, E.A., Mitzner, W.: Measurement of three-dimensional lung tree structures using computed tomography. J. Appl. Physiol. 79(5), 1687–1697 (1995)
12. Mayer, D., Bartz., D., Ley, S., Thust, S., Heussel, C.P., Kauczor, H.U., Straer, W.: Segmentation and virtual exploration of tracheobronchial trees. In 17th Int. Cong. and Exhibition Computer Aided Radiology and Surgery, pp. 35–40, London, UK (2003).
13. Chabat, F., Xiao-Peng, H., Hansell, D.M., Guang-Zhong, Y.: ERS transform for the automated detection of bronchial abnormalities on CT of the lungs. IEEE Trans. Med. Imag. 20(9), 942–952 (2001)
14. Busayarat, S., Zrimec, T.: Automatic detection of pulmonary arteries and assessment of bronchial dilatation in HRCT images of the lungs. In Int. ICSC Con. Computational Intelligence: Methods & Applications, pp. 1–6, Istanbul, Turkey, (2005).
15. Park, W., Hoffman, E.A., Sonka, M.: Segmentation of intrathoracic airway trees: a fuzzy logic approach. IEEE Trans. Med. Imag. 17(4), 489–497 (1998)
16. Rizi, F.Y., Ahmadian, A., Sahba, N., Tavakoli, V., Alirezaie, J., Fatemizadeh, E., Rezaie, N.: A Hybrid fuzzy based algorithm for 3D human airway segmentation. In 2nd International Conference on Bioinformatics and Biomedical Engineering, pp. 2295–2298, IEEE Press, New York (2008)
17. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Prentice Hall, New Jersey, USA (2007)
18. Dunn, J.: A Fuzzy Relative of the ISODATA Process and its Use in Detecting Compact, Well Separated Clusters. J. Cyber, 3, 32–57 (1974)
19. Ridler, T., Calvard S.: Picture thresholding using an iterative selection method. IEEE Trans. Syst. Man Cyb., 8, 630–632 (1978)
20. Extraction of Airways from CT 2009 `http://image.diku.dk/exact/index.php`
21. OpenMP Application Programming Interface `http://www.openmp.org`
22. Hasegawa, M., Nasuhara, Y., Onodera, Y., Makita, H., Nagai, K., Fuke, S., Ito, Y., Betsuyaku, T., Nishimura, M.: Airflow Limitation and Airway Dimensions in Chronic Obstructive Pulmonary Disease. Am. J. Respir. Crit. Care Med. 173(12), 1309–1315 (2006)
23. Rennard, S.I., Rodriguez-Rosin R., Huchon, G., Roche, N.: Clinical Management of Chronic Obstructive Pulmonary Disease. Informa Healthcare, UK (2007)

Robust Region Growing Based Intrathoracic Airway Tree Segmentation

Rômulo Pinho, Sten Luyckx, and Jan Sijbers

University of Antwerp, Physics Department, VisionLab, Belgium
{romulo.pinho; jan.sijbers}@ua.ac.be
{sten.luyckx}@student.ua.ac.be

Abstract. In this paper, we present a semi-automatic region growing algorithm to segment the intrathoracic airway tree from 3-d CT images. A common problem with region growing is leakage. In order to limit leakage, our method bounds the segmentation using cylinders of adaptive orientation and dimensions. The leaks are detected based on anatomical information of the airways and an algorithm to avoid them is proposed. We also present an algorithm to automatically select a seed point for the segmentation. The method was tested on a dataset of 40 patients and results were quantitatively evaluated based on ground truth data.

1 Introduction

Airway tree segmentation is the process of identifying and extracting from volumetric medical images the structures of the respiratory system that lead the air into the lungs. With the result of the segmentation, doctors and researchers can make measurements, check for abnormalities and generally be assisted in diagnosing diseases in the respiratory system.

In this work, we concentrate on the segmentation of the lower airway tree, namely the trachea, bronchi and bronchioli. Due to the natural complexity of the airways, with several branching levels, and noise or other artefacts present in the image, the segmentation is far from trivial. A common method to solve the problem is region growing [1], and semi- and fully-automated region growing algorithms have been used to segment the airways [2–4]. In this process, the user provides one or more seed points inside the airway structure. From these points, a region is grown by recursively aggregating voxels that pass a certain test of similarity. Common similarity tests check differences in intensity between neighbouring voxels.

One common problem of region growing algorithms is leakage. In the case of the airway tree segmentation, a thin wall separates the structure from neighbouring organs and air inside the lungs. Noise or other artefacts can create holes in this wall and, since the airway lumen and the lung interior have similar voxel intensities, the entire lung can be aggregated to the region. Another problem specific to airway tree segmentation is the early collapse of branches. In this case, the growing process stops too early, resulting in only partially segmented branches.

Tschirren *et al.* proposed an algorithm that takes advantage of the fact that the airway tree is a hierarchical combination of cylindrically shaped objects [5]. In their algorithm, cylinders of adaptive radius and orientation bound the segmentation, facilitating the process of leak detection. Later, Pinho *et al.* proposed improvements to [5] and introduced new ways of detecting leaks using anatomical, instead of pure image features [6]. In the present work, we build upon [6] and add the following contributions:

- propose a heuristic algorithm to automatically select a seed point inside the trachea, since chest CT scans often include the upper airways and other regions, complicating this task;
- use cylinders of adaptive height as well as adaptive radius in order to bound the segmentation;
- propose a new strategy to avoid leaks, by taking into account the fact that they grow through small holes on the edges of the structure being segmented.

The proposed method was evaluated with a dataset of 40 patients, subdivided into training and testing groups. Measures of number of branches, airway tree length, and leakage were taken in order to evaluate the method, by comparing it to ground truth data.

2 Method

We begin with a review of the method proposed in [5] and the ideas introduced in [6]. In [5], multiseeded fuzzy connectivity (MFC) [7] was used to segment the airways' walls and lumen in an iterative process which places adaptive cylinders (or ROIs) around the region to be grown. Airway walls and lumen compete for voxels based on an affinity value $\psi \in [0, 1]$ assigned as a function of voxel intensities. The ROIs bound the region growing and set limits to possible leaks. A leak detector assumes that leaks have a "spongy" structure, with many holes and tunnels. Once a leak is detected with a special morphological operator, the algorithm goes back to the previous step and repeats the segmentation, using what the authors called "directional affinity". This strategy avoids new leaks by assigning affinity values as a function of the intensity of a voxel and of its neighbours lying in the direction of the ROI. Airway branching is detected by computing the skeleton of a region within an ROI, using distance transforms. The branches of the skeleton, their spatial orientation, and the intersections between the region and the borders of the ROI determine the radius and orientation of the ROIs of the next step. The heights of ROIs may change if the segmentation stops exactly at a branching point. This process continues until no more voxels are aggregated.

In [6], the authors proposed improvements to the above algorithm. First, execution speed was increased with a simplification of the skeleton computation: instead of computing the real skeleton, an approximation was obtained by directly linking the centres of gravity of intersections between a region and their respective ROI. For intersections occurring on the side and upper borders of

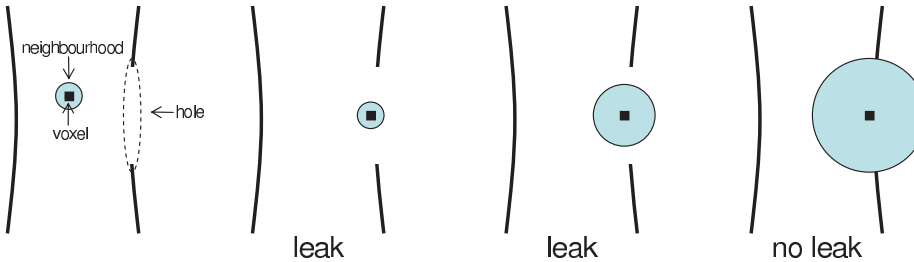


Fig. 1. Avoiding leaks. The segmentation is repeated with an increasing neighbourhood mask until no leaks are detected.

the ROIs, a global centre of gravity for the ROI was obtained and connected to the regions of intersection. This approximation is certainly not precise enough with respect to skeleton accuracy, but suffices for the estimation of the radii and orientations of the ROIs of the next step.

The second improvement dealt with the detection of leaks. Instead of the purely image based approach adopted in [5], the leak detection uses anatomical knowledge about the airways. For instance, the number of offspring branches from one level to the next is usually not larger than 3 or 4 and the radius of a branch is normally a decreasing function of its length and branching level. By checking the number of branches and their areas resulting from intersections with ROIs, leaks can be easily detected. Similar ideas were used in [8], with wavefront propagation algorithms, and in [9], with region growing.

In order to further improve [6] by detecting more airway branches and reaching deeper into the lungs, we propose to extend the region growing with two new ideas. The first is to use ROIs of adaptive height instead of only adaptive radius, such that $h_l = H\rho^{(l-1)}$ and $h_l \geq H_{min}$, where $l \geq 1$ is the current branching level, h_l is the height of the associated cylinder, in millimetres, H is the default initial height, $\rho \in [0, 1]$ is the height change ratio, and H_{min} is the minimum height limit. By doing this, short branches at higher branching levels, which otherwise would not intersect an ROI, can be detected.

Secondly, whenever the segmentation within an ROI is repeated due to a leak, each candidate voxel and its neighbourhood within a mask are analysed. Only if the voxel and all its non-visited neighbours pass the similarity test (voxel intensity lower than a threshold T in our case) is the voxel aggregated to the region. Each time the segmentation is repeated, a mask of higher radius is used, until no leaks are detected or a maximum number of attempts is reached. The reasoning behind this approach is that a leak always occurs due to the presence of holes on the boundaries of the region being segmented, so we are basically trying to discover their sizes. In contrast to the direction affinity adopted in [5], our neighbour affinity technique allows more possibilities of continuing with the segmentation, while trying to avoid leaks. The 2-d scheme of Fig. 1 illustrates the idea and **Algorithm 1** details the process.

Algorithm 1 region_grow(image, seed)

```
 1: /* intialization and computation of 1st ROI */
 2: while ¬ empty(roi_queue) do
 3:    r ← pop(roi_queue)
 4:    roi_region_grow(image, r)
 5:    has_leak ← detect_leak(r) {using anatomical information}
 6:    if has_leak ∧ (count_leak < max_count_leak) then
 7:       set_neighbour_affinity(mask[count_leak]) {set mask to avoid leak}
 8:       count_leak ← count_leak + 1
 9:       reset(image, r)
10:       push(roi_queue, r)
11:    else
12:       if has_leak then
13:          remove_leaking_branches(r) {may remove all branches}
14:       end if
15:       roi_list ← process_roi(r)
16:       for all r_i ∈ roi_list do
17:          push(roi_queue, r_i)
18:       end for
19:       count_leak ← 0
20:       set_neighbour_affinity(null) {switch neighbour affinity off}
21:    end if
22: end while
```

In the algorithm, Step 1 comprises a sequence of steps to compute the first ROI, using the given image and seed point, and push it onto an ROI-queue. After growing a region within the ROI at the front of the queue, Step 5 detects leaks using anatomical information. A leak is identified when the ROI splits into more than 5 regions for levels 1 through 4 and into more than 3 regions for higher levels. In addition, a leak is also identified when the area from a branch to its children increases by a factor $f \geq 2$. The regions of intersection between the ROI and the region grown corresponding to leaks are put in a list. If a leak is detected, Step 7 switches neighbour affinity on by providing the next neighbourhood mask to be used in the similarity tests. The ROI is reset and pushed back onto the queue in Steps 9 and 10, respectively. If leaks are still present after trying all masks, Step 13 removes from the list obtained in Step 5 the corresponding branches. Step 15 processes the remaining branches of the current ROI and returns a list of ROIs for the next iteration, which are pushed onto the queue. Finally, Step 20 switches neighbour affinity off and the process restarts.

Seed Point Selection As mentioned previously, region growing algorithms need one or more seed points to mark the start of the segmentation. Algorithms that automate the seed point selection for the segmentation of the airways typically detect a circular region near the centre of a slice of the image volume. The region is supposed to correspond to the trachea, and the seed point is taken as

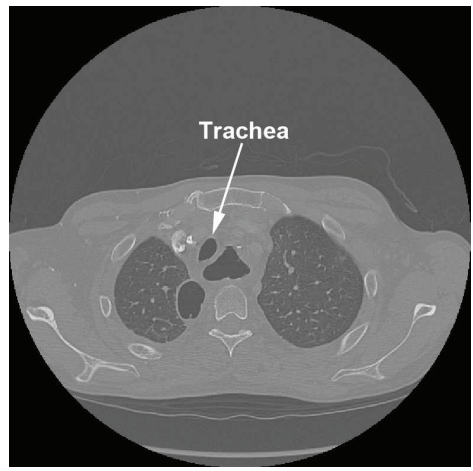


Fig. 2. A situation where other structures may mislead the detection of the trachea in an axial slice of the CT scan.

its centre of gravity. This process may fail if the chosen slice contains misleading regions (e.g., if the CT scan contains parts of the upper airways) or does not contain the trachea at all. Fig. 2 illustrates the former case with a slice containing the trachea, the oesophagus, and a tumour. In the present work, we propose a more robust method to automatically select a seed point inside the trachea.

For one axial slice $i = 1 \dots N$ of the image volume, the method works as in **Algorithm 2**. Let us use the *threshold below* operation to turn all voxels with intensity below a certain threshold to white and the rest to black [10]. Step 1 thus finds the best threshold to segment the air in the image, which includes the areas inside the lungs and airway lumen, using Otsu's method [11]. Step 2 applies a masked, morphological closing operation to the slice in order to fill all holes. Step 3 identifies 8-connected regions in the resulting image and labels them. Step 6 removes noise, i.e., all regions with size $s \leq S_{min}$ pixels. Step 10 takes care of eliminating narrowed regions, i.e., with excentricity $e > e_{max}$, and steps 12 through 15 identify the region of the slice with highest excentricity, R_{e_i}.

After these steps, a number of regions may be left in each slice. These regions comprise the trachea and areas corresponding to air outside the lungs, the upper airways, the lungs, etc. The challenge is then to choose the slice containing only the trachea or at least to correctly identify it when other structures are present. For this, we minimise a function of several parameters, in order to favour:

- slices with fewer regions, since, in general, the upper part of the trachea tends to appear alone in the image;
- slices that maximise e, of R_{e_i}, given that the upper part of the trachea, just below the infraglotic cavity, tends to be elliptical;
- slices in which the major axis of R_{e_i} is aligned with the sagittal plane;

Algorithm 2 find_trachea(slice$_i$)

```
threshold(slice_i) {segment air}
close(slice_i)
regions_i ← label(slice_i)
for all R_j ∈ regions_i do
    if size(R_j) < S_min then
        remove(regions_i, R_j) {remove noise}
    else
        e ← excentricity(R_j)
        if e > e_max then
            remove(regions_i, R_j) {remove narrow regions}
        else
            if e > max_e then
                max_e ← e
                R_ei ← R_j
            end if
        end if
    end if
end for
```

- slices with lower indexes, since the search is for the top of the trachea (assuming slice 0 coincides with the top position of the CT scan);
- slices in which R_{e_i} is small, which avoids confusion with the lungs;
- slices in which R_{e_i} maximises the area of the ellipse, so that only "regularly" shaped ellipses are chosen.

We therefore define the minimisation as

$$\arg\min_{i=1..N} f(i, e_i, a_i, s_i, r_i) = n_{r_i}\left(\frac{i}{N} + \frac{s_i}{S} + (1 - e_i) + a_i + (1 - r_i)\right), \quad (1)$$

where $\{e_i, a_i, r_i \in [0, 1]\}$. In this equation, i is the slice index, $n_{r_i} \geq 1$ is the number of regions of the slice, s_i is the size of R_{e_i} in pixels, with S being a maximum size threshold, e_i is the excentricity of R_{e_i}, a_i is the angle between R_{e_i}'s major axis and the sagittal direction, and r_i is a measure of area maximisation. The latter is computed by taking the ratio between the number of pixels of R_{e_i} and the area of its corresponding ellipse. Lastly, the selected seed point is the centre of gravity of the R_{e_i} that minimises $f(\cdot)$.

3 Experiments

As stated in Section 1, the proposed method was tested with a dataset of 40 patients, provided as part of the workshop and airway segmentation challenge *EXACT09: Extraction of Airways from CT 2009*. The data was subdivided into one training and one testing group, each with 20 patients, numbered CASE01...CASE20 and CASE21...CASE40, respectively. The segmentation was evaluated by a team of trained observers. The aim of the workshop

was to compare the performance of different algorithms. For this purpose, a ground truth was constructed from all submitted segmentations and all submissions were evaluated with respect to this ground truth.

The objective of the experiments was to check, for the testing group, how many branches were detected, the segmented tree length and the amount of leakage. The following measures were used to compare the submitted results:

- Branch count: number of branches detected.
- Branch detected: the fraction of branches that were detected with respect to the branches present in the ground truth.
- Tree length: the sum of the length of the centre lines of all correctly detected branches.
- Tree length detected: the fraction of tree length that was detected correctly, relative to the tree length of the ground truth.
- Leakage count: the number of unconnected groups of "correct" regions that are neighbours of a "wrong" region.
- Leakage volume: the volume of regions that are wrongly detected.
- False positive rate: the fraction of the volume of regions that are detected wrongly relative to the volume of all detected regions.

The trachea was excluded from the branch length and branch count related measurements. For the voxel based measures of leakage, both trachea and main bronchi were excluded. Furthermore, the exact airway shape and dimensions were not taken into account.

We implemented algorithms 1 and 2 in C++, and the programs were executed on an Intel® Core™ 2 Quad CPU, at 2.4 GHz, with 8GB of RAM, running under Windows Vista™ Ultimate 64-bits. The region growing algorithm used a single threshold value $T = -800HU$ for the whole airway tree and did not employ multiseeded connectivity, as opposed to [5], since we only segmented the airway lumen, not the walls. In addition, intensities of candidate voxels were averaged within a 6-connected neighbourhood to reduce noise artefacts. The parameters ρ and H_{min} were primarily chosen empirically for the training group, but adjustments were necessary during the experiments with the test set. Eventually, $\rho = 0.85$ provided the best results except for CASE32, for which it was set to $\rho = 0.75$, with $H_{min} = 2$mm in all cases. For the neighbour affinity, we used spherical and cubic masks with radii from 1 to 7 voxels, defining, in this order, 6, 18, 26, 92, 124, 342, 728, 1330, and 2197-neighbourhoods. These masks remained unchanged during the experiments with the test set, but new masks were added until the results for the training group were, at least visually, acceptable. With respect to the seed point selection, we used in **Algorithm 2** $S_{min} = 250$ pixels and $e_{max} = 0.75$, and, in Eq. (1), S was equal to the number of pixels of the slice and $N = \min(300, N_a)$ slices, where N_a is the number of axial slices of the image volume. Again, these values were empirically chosen for the training group, but remained unchanged with the test set.

3.1 Results

The results obtained with the region growing algorithm applied to the testing group[1] can be seen in Table 1. The main difficulty in the segmentation of the airways is to find the balance between the number of segments detected and leakage. In general, it is very difficult to increase the former without allowing the latter to increase as well. Our approach thus remained on the conservative side in terms of branch count and reach, but mostly with low leakage count.

Table 1. Evaluation measures for the twenty cases in the test set.

	Branch count	Branch detected (%)	Tree length (cm)	Tree length detected (%)	Leakage count	Leakage volume (mm^3)	False positive rate (%)
CASE21	69	34.7	39.4	35.7	0	0.0	0.00
CASE22	132	34.1	86.4	26.1	7	160.0	1.14
CASE23	89	31.3	56.6	21.7	6	56.1	0.52
CASE24	69	37.1	56.0	34.4	14	277.2	1.66
CASE25	76	32.5	58.5	23.2	8	557.8	3.27
CASE26	35	43.8	24.3	37.0	0	0.0	0.00
CASE27	36	35.6	25.9	31.9	0	0.0	0.00
CASE28	53	43.1	35.2	32.1	1	473.7	7.60
CASE29	73	39.7	46.9	34.0	4	27.6	0.40
CASE30	47	24.1	33.2	21.7	0	0.0	0.00
CASE31	61	28.5	39.0	22.2	7	578.0	6.98
CASE32	64	27.5	46.6	21.4	2	1740.7	14.34
CASE33	70	41.7	50.2	34.2	5	670.3	11.25
CASE34	140	30.6	85.4	23.9	10	2407.9	12.70
CASE35	95	27.6	61.1	19.8	3	39.7	0.32
CASE36	83	22.8	69.7	16.9	0	0.0	0.00
CASE37	67	36.2	52.3	29.4	2	105.9	1.11
CASE38	28	28.6	23.5	35.3	0	0.0	0.00
CASE39	109	21.0	84.8	20.7	2	93.5	1.04
CASE40	88	22.6	63.6	16.4	13	1420.6	10.24
Mean	74.2	32.1	51.9	26.9	4.2	430.4	3.63
Std. dev.	29.5	6.9	19.6	6.9	4.4	672.3	4.92
Min	28	21.0	23.5	16.4	0	0.0	0.00
1st quartile	53	27.5	35.2	21.4	0	0.0	0.00
Median	70	31.9	51.3	25.0	3	99.7	1.07
3rd quartile	95	39.7	69.7	34.4	8	670.3	10.24
Max	140	43.8	86.4	37.0	14	2407.9	14.34

Further improvements to the proposed method include more robust algorithms to detect leaks and avoid them. One characteristic of the neighbour affinity we adopted is the fact that the resulting segmentation will become thinner as

[1] Provided by the organisers of the workshop.

the neighbourhood mask increases, but this can be corrected with local dilation operations. Another improvement is the use of adaptive intensity thresholds, employed in, e.g., [9, 3]. In fact, we have already tried this approach and produced some primary results. We observed that although the number of branches may increase considerably, so may the number of leaks. As a consequence, this technique must be coupled with efficient leak detection and removal.

With respect to the automatic seed point selection, the proposed algorithm performed very well in all cases. The selected point was always located inside the trachea, at the top. In very few situations, however, the point was set at a lower location. This happened when the trachea had an almost circular shape along all or nearly all of its length. Since we favoured elliptical regions, such a shape happened to appear at slices with higher indexes. Given that the trachea was not considered in these experiments, it was not a problem, but adjustments to the function of Eq. (1) may still be necessary.

Table 2 presents the execution times of the proposed algorithms applied to the testing group. For the region growing, execution time is naturally an increasing function of the number of detected branches, but all executions ran in less than 1 minute, with half of them below 3 seconds and 75% below 8 seconds. The seed point selection, in turn, showed less varying execution times, mostly because $N = 300$ slices for all cases. The differences between cases lay mainly in the complexity of each slice processed by the algorithm. Finally, Fig. 3 presents the segmentation results for 2 patients.

4 Conclusions

In this work, we presented a semi-automatic region growing method for the segmentation of the intrathoracic airways from tomographic scans. The method uses cylinders (or ROIs) of adaptive orientation and dimensions to bound the segmentation. The role of these ROIs is to set a limit to leaks, a common problem with region growing algorithms, and to allow them to be more easily detected. Our approach uses anatomical information about the airways in order to detect the leaks and we proposed a novel algorithm to avoid new leaks once they are detected. We also proposed a heuristic algorithm to automatically select a seed point at the top of the trachea, which is later provided to the region growing algorithm. The method was tested on a dataset of 40 patients, and remained on the conservative side in terms of branch detection, but with a low number of leaks in most cases.

References

1. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Addison-Wesley Longman Publishing Co., Inc. (2001)
2. Chiplunkar, R., Reinhardt, J.M., Hoffman, E.A.: Segmentation and quantitation of the primary human airway tree. In: Proc. SPIE Medical Imaging. Volume 3033. (1997) 403–414

Table 2. Execution times of the algorithms applied to the test set.

	Region growing (secs)	Seed selection (secs)	Total time (secs)
CASE21	1.98	82.9	84.88
CASE22	6.85	79.6	86.45
CASE23	15.88	64.2	80.08
CASE24	15.02	63.2	78.22
CASE25	31.79	60	91.79
CASE26	1.59	68.8	70.39
CASE27	1.22	57.2	58.42
CASE28	1.83	61	62.83
CASE29	1.67	61.4	63.07
CASE30	0.97	61.2	62.17
CASE31	3.51	64.2	67.71
CASE32	25.74	66.8	92.54
CASE33	5.46	65.6	71.06
CASE34	44.62	57.1	101.72
CASE35	3.6	59.4	63
CASE36	1.62	69	70.62
CASE37	5.3	59.2	64.5
CASE38	2.79	58	60.79
CASE39	1.31	76.5	77.81
CASE40	2.22	70.2	72.42
Mean	8.75	65.27	74.02
Std. dev.	12.08	7.38	12.27
Min	0.97	57.1	58.42
1st quartile	1.66	59.85	63.05
Median	3.15	63.7	70.84
3rd quartile	8.89	68.85	81.28
Max	44.62	82.9	101.72

3. Kiraly, A.P., Higgins, W.E., McLennan, G., Hoffman, E.A., Reinhardt, J.M.: Three-dimensional human airway segmentation methods for clinical virtual bronchoscopy. Academic Radiology **9** (2002) 1153 – 1168
4. Law, T.Y., Heng, P.: Automated extraction of bronchus from 3d CT images of lung based on genetic algorithm and 3d region growing. Volume 3979. (2000) 906–916
5. Tschirren, J., Hoffman, E., McLennan, G., Sonka, M.: Intrathoracic airway trees: segmentation and airway morphology analysis from low-dose CT scans. Medical Imaging, IEEE Transactions on **24** (2005) 1529– 1539
6. Pinho, R., Sijbers, J., Vos, W.: Efficient approaches to intrathoracic airway tree segmentations. In: Proc. IEEE EMBS Benelux Symposium. Volume 2. (2006) 151–154
7. Udupa, J.K., Samarasekera, S.: Fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation. Graphical Models and Image Processing **58** (1996) 246 – 261

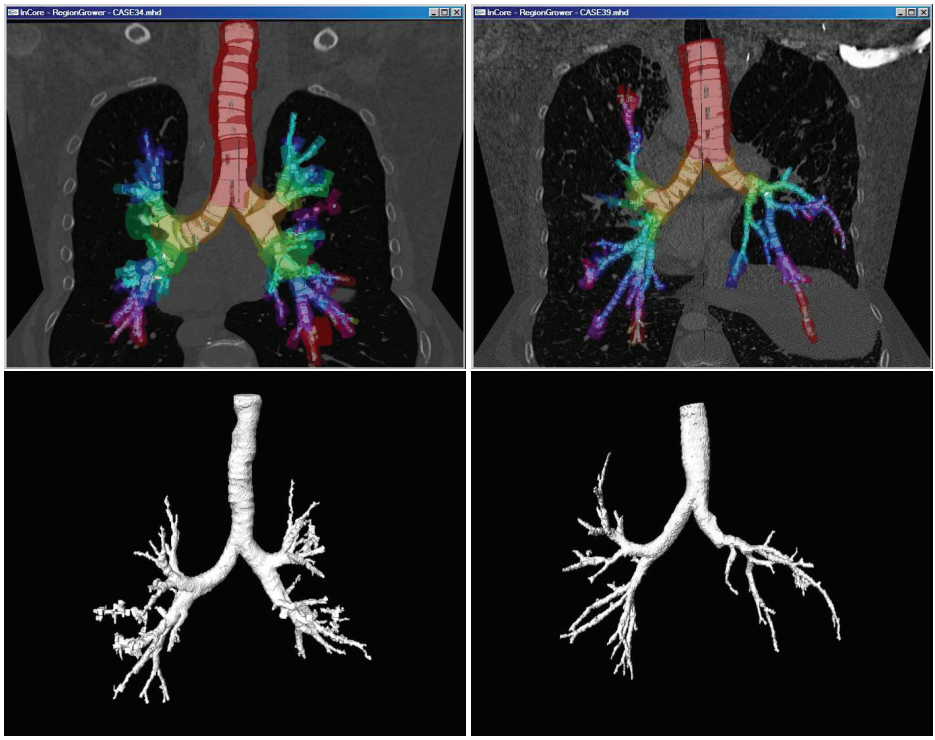


Fig. 3. Segmentations with ROIs coloured per level (T) and respective 3-d reconstructions (B) for CASE34 (L) and CASE39 (R), respectively.

8. Schlathoelter, T., Lorenz, C., Carlsen, I.C., Renisch, S., Deschamps, T.: Simultaneous segmentation and tree reconstruction of the airways for virtual bronchoscopy. In: Proc. SPIE Medical Imaging. Volume 4684. (2002) 103–113
9. van Ginneken, B., Baggerman, W., van Rikxoort, E.M.: Robust segmentation and anatomical labeling of the airway tree from thoracic CT scans. In: MICCAI. (2008) 219–226
10. Shapiro, L.G., Stockman, G.C.: Computer Vision. Prentice Hall (2002)
11. Otsu, N.: A threshold selection method from gray-level histograms. Systems, Man and Cybernetics, IEEE Transactions on **9** (1979) 62–66

Adaptive Branch Tracing and Image Sharpening for Airway Tree Extraction in 3-D Chest CT

Marco Feuerstein[1], Takayuki Kitasaka[2,3], Kensaku Mori[1,3]

[1] Graduate School of Information Science, Nagoya University, Japan,
fmarco@suenaga.m.is.nagoya-u.ac.jp,
[2] Faculty of Information Science, Aichi Institute of Technology, Japan,
[3] MEXT Innovation Center for Preventive Medical Engineering, Nagoya University, Japan.

Abstract. The automated extraction of the airway tree from 3-D chest CT data can greatly reduce the workload of physicians during diagnosis (e.g. quantification of airway morphology) and treatment (computer-aided bronchoscopy) of lung disease. This paper presents a method to automatically extract the airways driven by a sharpening filter, which enhances the branch edges in the input image based on the Laplacian of Gaussian, and adaptive cuboidal volumes of interest that an adaptive region growing algorithm uses to trace the airway tree. The method was trained on 20 data sets and evaluated on another 20 data sets from various scanners, using a wide range of acquisition and reconstruction parameters, including low dose scans. Compared to other state-of-the-art methods, our algorithm features the highest detection and extraction rates of bronchial branches. Future research needs to focus on the development of a method for automatic leakage detection.

1 Introduction

During the last century, from the first X-ray images taken 1895 and the advent of 3-D computed tomography (CT) in the seventies, medical imaging technologies greatly advanced. This went along with the generation of high-resolution data sets allowing more accurate as well as reliable diagnosis and treatment of diseases. However, as physicians diagnosing a patient need to handle this vast amount of data in a timely manner, various researchers and companies also started the development of methods for automated processing and analysis of patient data to allow a faster computer-aided diagnosis.

One example is the automatic extraction of the tracheobronchial (airway) tree from chest CT data to reduce the workload of physicians during diagnosis (e.g. quantification of airway morphology) and treatment (computer-aided bronchoscopy) of lung disease. This is a very active field of research, which started to evolve in about 1995 [1]. Sluimer et al. provide a short overview of methods developed during the first 10 years of research [2]. As all methods have to face problems highly related to image quality and acquisition such as low-dose scanning, noise, partial volume effects, and cardiac motion, and patient-specific

aspects such as pathology, implants, and mucus, the perfect method is yet to be developed. Hence, many more recent works can be found in the literature addressing one or more of these problems, e.g. [3–6].

We here present an enhancement of our previously developed adaptive airway tree extraction method [7, 8], which can work on noisy and low-dose scans, is faster than our initial method, and was evaluated on a large number of datasets.

2 Method

As our method requires input data with reasonably low noise and a seed point in the trachea to start the airway extraction, we are first performing a couple of preprocessing steps, as described in section 2.1. Section 2.2 describes the core of our algorithm. Basically, it first reconstructs a small volume of interest (VOI) around the current branch of the airway tree (starting at the trachea), which is automatically adapted to the size of the branch. Within the VOI, first the edges of the airways are enhanced by a sharpening filter, followed by region growing using an adaptive threshold to take care of obvious leakages and a routine to detect bifurcations or trifurcations of the tree. Each bifurcation or trifurcation is followed by generating more volumes of interest and repeating the process, until no more furcations can be detected.

2.1 Preprocessing

Seed Point Search As our algorithm depends on a reliable seed point inside the trachea, we first determine air voxels within the body region of the patient, whose intensities are lower than about -500 Hounsfield units (HU). Based on a method similar to [9], we then extract the lung and its bounding box. Within the most superior axial slice of the bounding box, we automatically determine the trachea, which is of circular shape, specific size (i.e. a diameter of less than about 5 cm), approximately in the center of the bounding box along the mediolateral axis, and in between the center and the anterior of the bounding box along the anteroposterior axis. It is important to use this particular slice to start the search and not another more superior slice e.g. in the oral cavity, as the air region there is not shaped as the trachea and sometimes disconnected from the rest of the airways. From all trachea voxels extracted in the axial slice we simply choose their centroid as our seed point for airway extraction.

Image Smoothing Our algorithm also requires image data that contains a reasonably low amount of noise, which usually does not hold for certain scan types (e.g. low dose scans) and reconstruction parameters (e.g. very sharp convolution kernels). We therefore analyze image noise by computing the mean of the gradient magnitude of all air voxels identified previously inside the bounding box of the lung. Depending on the level of image noise, we perform n image smoothing iterations. During smoothing, we utilize a modified curvature diffusion equation [10], which is a level-set equivalent of the anisotropic diffusion equation proposed

in [11]. However, it preserves and enhances airway tree edges better than conventional anisotropic diffusion. Using default parameters, i.e. a time step of 0.0625 in the computation of the level set evolution and a value of conductance of 3.0, we alter the number of iterations n depending on the mean gradient magnitude $\overline{\nabla}$:

$$n = \begin{cases} 0 & \text{if } \overline{\nabla} \leq 100 \\ 1 & \text{if } 100 < \overline{\nabla} \leq 150 \\ 2 & \text{if } 150 < \overline{\nabla} \leq 200 \\ 3 & \text{otherwise} \end{cases} \tag{1}$$

2.2 Airway Tree Extraction

An airway tree is extracted by tracing its structure branch by branch using VOIs of adaptive size and starting at the trachea. We here explain the procedures briefly. More details can be found in [7, 8, 12].

Initial Threshold Selection First, we obtain an initial threshold for airway tree extraction, T_{ini}, by performing a region growing based method [1], which extracts the largest region whose number of voxels is smaller than a predefined value (e.g. one tenth of the lung volume) starting at our previously determined seed point.

VOI Setting We define a VOI as a cuboid with eight vertices $\boldsymbol{V}_1$ - $\boldsymbol{V}_8$ and two points $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$, which are the centers of the two faces S_1 and S_2 (see Fig. 1). The size of a VOI is denoted by the width W, height H, and depth D. The z-axis of a VOI indicates the running direction of a bronchial branch. Hereafter, the image circumscribed by a VOI is used to calculate local features of the bronchus and is called *VOI image*.

Local Image Enhancement Whenever a new VOI image is generated, it is first enhanced by performing a sharpening filter based on the Laplacian of Gaussian [7]. Using the sharpening filter, the intensity of the bronchial wall becomes higher while that of the lumen is weakened. This is to reinforce bronchial features that were lost due to partial volume effects.

In detail, a VOI image $\boldsymbol{F}$ is enhanced to $\boldsymbol{F}_{\text{sharpened}}$ by

$$\boldsymbol{F}_{\text{sharpened}} = \boldsymbol{F} - \beta \cdot \frac{LoG(\boldsymbol{F}) + L'oG(\boldsymbol{F})}{2} \tag{2}$$

where LoG is the Laplacian of Gaussian, $L'oG$ is a modified version of it, and β determines the degree of sharpening. The $L'oG$ filter is using the same Gaussian convolution kernel as the LoG filter, but its Laplacian convolution kernel ignores any voxels being greater than the center voxel of the kernel.

The standard Laplacian filter gives negative values for voxels whose neighbors are of lower intensities and positive values for neighbors of higher intensities. This is, if airway branches are clearly separable from surrounding tissue, bronchial

wall voxels will tend to give negative values, while bronchial lumen voxels will tend to give positive values, and thus airway branches are enhanced. However, if the bronchial wall is adjacent to tissue of higher intensity than the wall itself, the standard *LoG* filter can provoke the contrary result, weakening the wall. To counteract this effect, we introduce the modified Laplacian of Gaussian, whose output is always negative and therefore never weakens the bronchial wall. We equally weight *LoG* and $L'oG$.

In our current implementation, we set the scale σ_L of the Laplacian kernel to

$$\sigma_L = \min\left(3, \lfloor 15 \cdot reso_{\min}/r \rfloor\right) \tag{3}$$

and the scale σ_G of the Gaussian kernel to

$$\sigma_G = reso_{\min} \tag{4}$$

where $reso_{\min}$ is the minimum voxel resolution of the VOI image and r is the radius of the currently processed branch (see below for details on how to obtain r). However, in the future we will unify the sizes of the convolution kernels, so they are both dependent on the radius of the current branch.

VOI for Trachea We set the VOI for the trachea, which is the root of the airway tree, based on the seed point detected above. First, the orientation of the coordinate system of the VOI is set to be equal to that of the CT image. The center point $\boldsymbol{P}_1$ of the VOI is translated so as to coincide with the seed point. The VOI is of cuboidal shape and its edges are twice as long as the diameter of the trachea region intersecting with the most superior axial slice of the bounding box of the lung. The bronchial region inside the VOI image is then extracted by a region growing with threshold T, which is initialized by $T_{ini} + 100$[HU].

Then, the VOI is iteratively adjusted in z direction using a binary search [12]. This is, the VOI is extended or shortened while halving the search interval of the binary search, until we find a furcation or the end of a branch (see below).

Identification of Furcation After each iteration, we check the number of connected components, N_c, on the VOI surface (except S_1) to find out whether the current branch furcates or not. Each connected component is denoted by C_i $(1 \leq i \leq N_c)$. Depending on the number of connected components, we iteratively execute one of the following cases:

- $\mathbf{N_c = 0}$: If the binary search interval is still greater than $reso_{\min}$, we shorten the VOI. If it reaches $reso_{\min}$, we terminate tracing.
- $\mathbf{N_c = 1}$: The VOI is extended and the bronchial region is re-segmented using region growing thresholded at T.
- $\mathbf{N_c = 2}$ or $\mathbf{N_c = 3}$: We found a furcation. If the binary search interval is still greater than $reso_{\min}$, we shorten the VOI. If it is equal to $reso_{\min}$, we terminate tracing for the current branch and calculate its final dimensions. Its thickness is measured by averaging the extracted bronchial region on each slice of the VOI image along the z-axis of the VOI. The radius of the

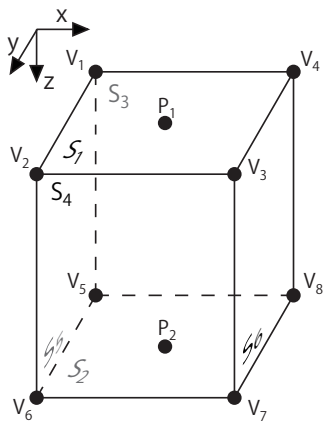


Fig. 1: Definition of VOI. It has eight vertices $\boldsymbol{V}_1$ - $\boldsymbol{V}_8$ and two points $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ that are the centers of the faces S_1 and S_2. The VOI size is $W \times H \times D$.

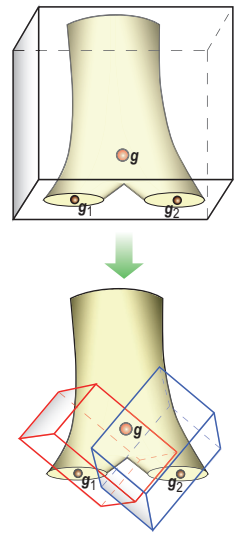


Fig. 2: Generation of child VOIs. Each child VOI is oriented along the line from a detected furcation point $\boldsymbol{g}$ to the center of gravity $\boldsymbol{g}_i$ of child branches. Width and height of a child VOI are set to be twice as large as the radius of the child branch.

branch, r, is calculated from this thickness. The furcation point, $\boldsymbol{g}$, is then determined as the gravity center of the bronchial region on the (D-r)-*th* slice. The gravity center of each C_i, $\boldsymbol{g}_i$, is also calculated. Then, child VOIs are generated based on $\boldsymbol{g}$, $\boldsymbol{g}_i$, and r, as described in the next paragraph.

– $\mathbf{N_c > 3}$: If the binary search interval is still greater than $reso_{\min}$, we shorten the VOI. If it is equal to $reso_{\min}$, the threshold T is reduced by ΔT and the bronchial region is updated.

Generation of Child VOIs If a bifurcation or trifurcation appears, we generate child VOIs for each C_i to further trace the airway branches (Fig. 2). P_1 of a child VOI for C_i is set to the furcation point $\boldsymbol{g}$, while P_2 is set to $\boldsymbol{g}_i$. That is, the VOI is oriented from the furcation point to the child branch. Width and height of a child VOI are set to be twice as big as the radius of the child branch, which is calculated from the region C_i. To prevent tracing from growing into sibling branches, we set levees L_j among VOIs as shown in Fig. 3. No voxels beyond these levees are extracted. Then, the bronchial region in a child VOI is updated. If the radius of a branch is smaller than a predefined value T_r, then the resolution of the current VOI image is doubled by using tricubic interpolation. The bronchial

region is updated again and we repeat the iterative process described before to check whether a furcation appears.

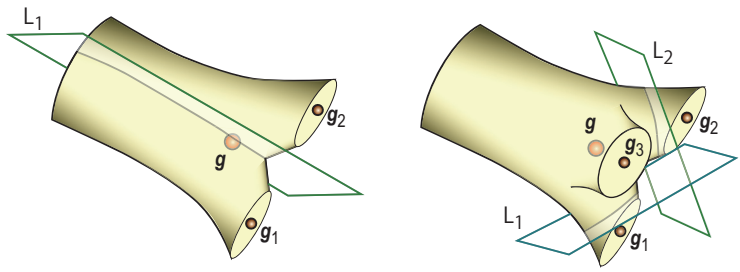


Fig. 3: Levees to prevent tracing from growing into sibling branches. They are set among newly created child VOIs for the cases of (left) bifurcation and (right) trifurcation, respectively.

Reconstruction of Airway Tree After finishing all processes, we reconstruct the airway tree by projecting bronchial regions inside all VOI images into the original CT image. In addition, the tree structure is extracted by connecting P_1 and $\boldsymbol{g}$ of each VOI.

3 Results

Our method was centrally evaluated by a team of trained observers within the scope of an airway tree segmentation challenge to compare the performance of different algorithms [13]. We applied our method to 40 CT data sets acquired by various scanners located in hospitals in Denmark, Germany, the Netherlands, USA, and Japan, using a wide range of acquisition and reconstruction parameters, including low dose scans. We used 20 data sets to train our method and 20 data sets for evaluation. The free parameters of our algorithm, as described in Section 2, were selected as $\Delta T = 4$, $T_r = 15 \cdot reso_{\min}$, and $\beta = 0.05$, using the training data.

To obtain correctly (true positive) and falsely (false positive) extracted airway tree voxels and branches and in order to compare our method to other algorithms, ground truth segmentations were generated from the extraction results of all algorithms. Therefore, for each algorithm and data set, following evaluation steps were performed:

1. Determination of all airway tree branches from the segmentation using a fast marching based algorithm similar to [14].
2. Visual evaluation of the airway tree branches using MPR (multiplanar reformation) and oblique views as well as reformatted views with straightened airway centerlines.

3. Labeling of each branch to be true positive or false positive, i.e. whether it is part of the airway tree or not, by at least two observers; its exact shape and dimensions are ignored.

The overall ground truth segmentation for each data set is then generated from all true positive branches of all algorithms. For evaluation, following seven measures were defined and computed:

- *Branch count:* Number of true positive branches with a centerline length of more than 1 mm.
- *Branch detected:* Percentage of the number of true positive branches divided by the number of ground truth branches.
- *Tree length:* Overall length of the centerlines of all true positive branches.
- *Tree length detected:* Percentage of the true positive tree length divided by the ground truth tree length.
- *Leakage count:* Number of unconnected true positive regions in a 26-connected neighborhood of all false positive voxels, i.e. within a one voxel-thick border around all false positive voxels. This measure indicates how difficult it is to manually separate leakages from true positive branches.
- *Leakage volume:* The volume of all false positive voxels.
- *False positive rate:* Percentage of the number of false positive voxels divided by the number of all detected voxels.

During evaluation, the trachea was excluded from any measures related to branch and tree length or count. For the leakage measures, the trachea as well as the left and right main bronchi were excluded.

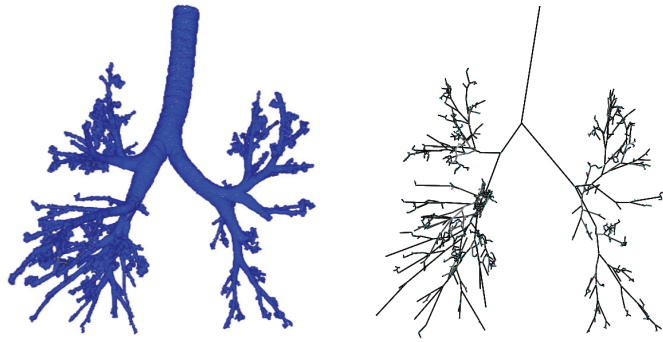
Table 1 summarizes the results of the evaluation of our method. Exemplary airway tree extractions depicting extreme cases of leakages and branch extractions are shown on Fig. 4. Depending on the size of the data set and the number of branches to extract, our method takes about 5 ± 3 minutes to process a case on a standard workstation with two 64-bit Quad-Core Intel Xeon 5355 processors.

4 Discussion

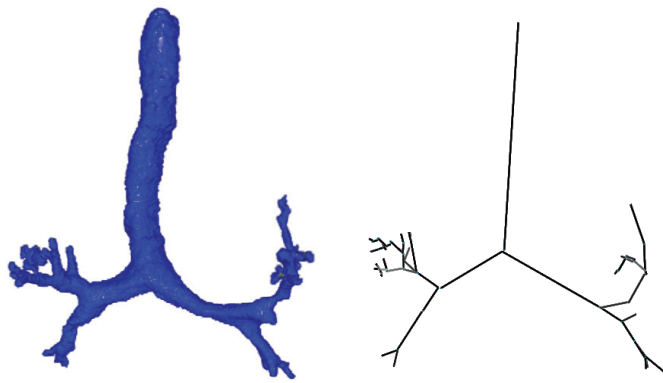
As can be seen in Tables 1 and 2, our method can extract the airway tree more completely than any method of the other teams. However, it is also obvious that our method currently produces a lot of leakages. All leakages occur in peripheral branches, where the bronchus walls are very thin and hence allow their interior to easily merge with the exterior (due to partial volume effects and insufficient image resolution). The lacking availability of a more appropriate stopping criterion or leakage detection of our algorithm attributes to the high number of false positive branches. Various solutions for this could be thought of. One reasonable approach would be to integrate simultaneous labeling of branches into our algorithm (similar to [14]), which could verify the currently extracted tree structure against an atlas or a database of ground truth trees and automatically detect branches that are heavily deviating from the ground truth data,

Table 1: Evaluation results for the twenty cases in the test set.

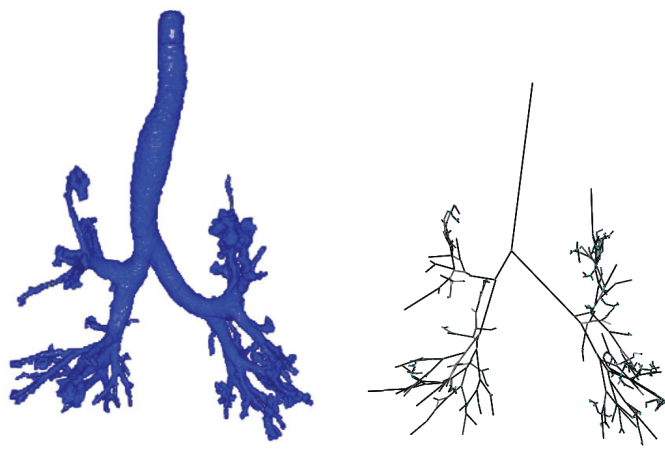
	Branch count	Branch detected (%)	Tree length (cm)	Tree length detected (%)	Leakage count	Leakage volume (mm^3)	False positive rate (%)
CASE21	104	52.3	58.4	52.8	9	1138.8	10.43
CASE22	292	75.5	233.6	70.7	103	7558.5	20.37
CASE23	242	85.2	198.4	76.3	92	5387.7	16.83
CASE24	164	88.2	140.3	86.3	50	6690.9	18.04
CASE25	210	89.7	226.3	89.8	33	4944.8	11.99
CASE26	42	52.5	30.6	46.5	2	92.6	1.52
CASE27	72	71.3	48.9	60.3	2	63.2	0.61
CASE28	116	94.3	97.4	88.8	13	2011.7	12.68
CASE29	137	74.5	99.8	72.3	21	4417.9	22.42
CASE30	157	80.5	117.6	77.0	24	10404.7	35.88
CASE31	187	87.4	158.7	90.4	30	11956.8	30.90
CASE32	181	77.7	166.5	76.4	34	8452.1	21.75
CASE33	138	82.1	111.7	75.9	30	1722.3	12.28
CASE34	339	74.0	248.7	69.5	61	5287.9	11.47
CASE35	301	87.5	270.2	87.3	70	7188.3	17.34
CASE36	265	72.8	274.6	66.6	7	809.2	2.43
CASE37	170	91.9	155.4	87.4	42	7966.4	24.28
CASE38	66	67.3	43.9	66.1	5	255.3	2.24
CASE39	248	47.7	202.6	49.5	34	5817.6	16.76
CASE40	305	78.4	290.9	75.2	47	10598.2	21.08
Mean	186.8	76.5	158.7	73.3	35.5	5138.2	15.56
Std. dev.	86.4	13.3	82.2	13.4	28.6	3754.5	9.52
Min	42	47.7	30.6	46.5	2	63.2	0.61
1st quartile	116	71.3	97.4	66.1	9	1138.8	10.43
Median	176	78.0	157.0	75.6	32	5337.8	16.79
3rd quartile	292	88.2	248.7	87.4	61	8452.1	22.42
Max	339	94.3	290.9	90.4	103	11956.8	35.88



(a) CASE22 (Highest number of leakages)

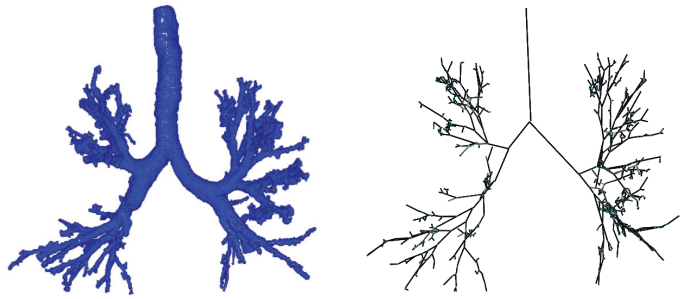


(b) CASE26 (Lowest number of detected branches)

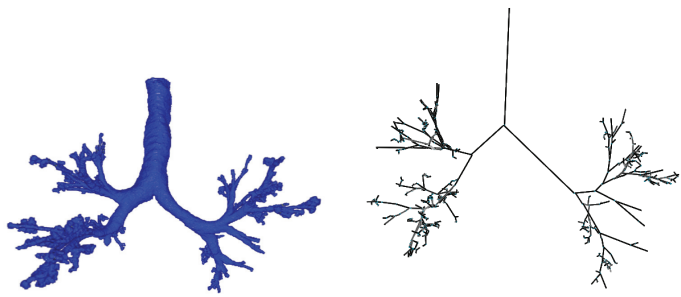


(c) CASE31 (Highest leakage volume)

Fig. 4: Exemplary results visualized by (left) volume rendering of the extracted airways and (right) the corresponding tree structure.



(d) CASE34 (Highest number of detected branches)



(e) CASE37 (Highest percentage of detected branches)

(f) CASE40 (Longest tree length)

Fig. 4: Exemplary results visualized by (left) volume rendering of the extracted airways and (right) the corresponding tree structure.

e.g. in terms of branch length, branch direction, or furcation patterns. Another more pragmatic solution would be to check for extracted regions not looking like branches, i.e. having a much bigger radius than their parent branches or not being of tubular shape. This could be performed in a post-processing step.

Table 2: Comparison of our results to all other teams participating in the challenge. Teams using semi-automated segmentation methods are marked by a *.

	Branch count	Branch detected (%)	Tree length (cm)	Tree length detected (%)	Leakage count	Leakage volume (mm^3)	False positive rate (%)
Team 1	91.1	43.5	64.6	36.4	2.5	152.3	1.27
Team 2	157.8	62.8	122.4	55.9	12.0	563.5	1.96
*Team 3	74.2	32.1	51.9	26.9	4.2	430.4	3.63
(Ours) Team 4	186.8	76.5	158.7	73.3	35.5	5138.2	15.56
Team 5	150.4	59.8	118.4	54.0	1.9	18.2	0.11
*Team 6	77.5	36.7	54.4	31.3	2.3	116.3	0.92
Team 7	146.8	57.9	125.2	55.2	6.5	576.6	2.44
*Team 8	71.5	30.9	52.0	26.9	0.9	126.8	1.75
Team 9	139.0	56.0	100.6	47.1	13.5	368.9	1.58
Team 10	79.3	32.4	57.8	28.1	0.4	14.3	0.11
*Team 11	93.5	41.7	65.7	34.5	1.9	39.2	0.41
Team 12	130.1	53.8	95.8	46.6	5.6	559.0	2.47
Team 13	152.1	63.0	122.4	58.4	5.0	372.4	1.44
Team 14	161.4	67.2	115.4	57.0	44.1	1873.4	7.27
*Team 15	148.7	63.1	119.2	58.9	10.4	158.8	1.19

Furthermore, it is worth mentioning that our algorithm as well as any other algorithms based on region growing, level sets, fast marching, and the like starting from a single seed point inside the trachea will fail to extract the airways, if any of the branches are interrupted by tumors. This issue needs to be tackled by the research community in their upcoming work.

5 Conclusion

We presented a fully automatic method for airway tree extraction in 3-D chest CT. After automatically finding a seed point in the trachea, it tracks and extracts the airway branches within adaptive cuboidal volumes of interest and by enhancing the input image utilizing a sharpening filter based on the Laplacian of Gaussian. As our current implementation does not include an automatic leakage detection method, we get false positive branches in the peripheral branches. Apart from that, our method is able to extract a very high number of airway branches, in fact more than any of the other 14 semi-automated and automatic methods evaluated within an airway extraction challenge.

Acknowledgments

The authors thank their colleagues for many helpful suggestions and advices. Parts of this research were supported by the Japan Society for the Promotion of Science (JSPS) postdoctoral fellowship program for foreign researchers, a Grant-In-Aid for Scientific Research from JSPS, the program of formation of innovation center for fusion of advanced technologies "Establishment of early preventing medical treatment based on medical-engineering for analysis and diagnosis" funded by the Ministry of Education (MEXT), and a Grant-In-Aid for Cancer Research from the Ministry of Health, Labour and Welfare.

References

1. Mori, K., Hasegawa, J., Toriwaki, J., Anno, H., Katada, K.: Automated extraction and visualization of bronchus from 3D CT images of lung. In: Proceedings of Computer Vision, Virtual Reality and Robotics in Medicine. (1995)
2. Sluimer, I., Schilham, A., Prokop, M., van Ginneken, B.: Computer analysis of computed tomography scans of the lung: A survey. IEEE Transactions on Medical Imaging **25**(4) (2006) 385–405
3. Tschirren, J., Hoffman, E., McLennan, G., Sonka, M.: Intrathoracic airway trees: Segmentation and airway morphology analysis from low-dose CT scans. IEEE Transactions on Medical Imaging **24**(12) (2005) 1529–1539
4. Graham, M.W., Gibbs, J.D., Higgins, W.E.: Robust system for human airway-tree segmentation. In: SPIE Medical Imaging. (2008)
5. Lo, P., de Bruijne, M.: Voxel classification based airway tree segmentation. In: SPIE Medical Imaging. (2008)
6. van Ginneken, B., Baggerman, W., van Rikxoort, E.M.: Robust segmentation and anatomical labeling of the airway tree from thoracic CT scans. In: MICCAI. (2008)
7. Kitasaka, T., Mori, K., Hasegawa, J., Toriwaki, J.: A method for extraction of bronchus regions from 3D chest X-ray CT images by analyzing structural features of the bronchus. FORMA **17**(4) (2002) 321–338
8. Kitasaka, T., Mori, K., Suenaga, Y., Hasegawa, J., Toriwaki, J.: A method for segmenting bronchial trees from 3D chest X-ray CT images. In: MICCAI. (2003)
9. Hu, S., Hoffman, E.A., Reinhardt, J.M.: Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. IEEE Transactions on Medical Imaging **20**(6) (2001) 490–498
10. Whitaker, R., Xue, X.: Variable-conductance, level-set curvature for image denoising. In: International Conference on Image Processing. Volume 3. (2001) 142–145
11. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence **12**(7) (1990) 629–639
12. Mori, K., Kitasaka, T., Akimoto, S., Ebe, K., Wada, T.: Endoscope insertion support system and endoscope insertion support method (2009)
13. Lo, P., van Ginneken, B., Reinhardt, J., de Bruijne, M.: Extraction of airways from CT (EXACT'09). In: Second International Workshop on Pulmonary Image Analysis. (2009)
14. Schlathölter, T., Lorenz, C., Carlsen, I.C., Renisch, S., Deschamps, T.: Simultaneous segmentation and tree reconstruction of the airways for virtual bronchoscopy. In: SPIE Medical Imaging. (2002)

Maximal Contrast Adaptive Region Growing for CT Airway Tree Segmentation

Carlos S. Mendoza *, Begoña Acha, and Carmen Serrano

Universidad de Sevilla
Av. de los Descubrimientos s/n
41092 Sevilla, Spain
{csanchez1,bacha,cserrano}@us.es

Abstract. In this paper we propose a fully self-assessed adaptive region growing airway segmentation algorithm. We rely on a standardized and self-assessed region-based approach to deal with varying imaging conditions. Initialization of the algorithm requires prior knowledge of trachea location. This can be provided either by manual seeding or by automatic trachea detection in upper airway tree image slices. The detection of the optimal parameters is managed internally using a measure of the varying contrast of the growing region. Extensive validation is provided for a set of 20 chest CT scans. Our method exhibits very low leakage into the lung parenchyma, so even though the smaller airways are not obtained from the region growing, our fully automatic technique can provide robust and accurate initialization for other methods.

Key words: CT, airways, segmentation, multi-tolerance, adaptive region growing, median filter

1 Introduction

CT scans are frequently used for pulmonary disorder assessment [1]. Pathologies that could affect sufficient lung function include tumors, pulmonary embolism, atelectasis, pneumonia, emphysema, asthma, bronchiectasis, and many others. Certain lung diseases can be diagnosed based on airway wall thickness measurements, diameter, branching geometry and rate of tapering. CT is currently the only readily accessible, relatively noninvasive technique that is capable of providing airway tree quantitative structural data in vivo [2].

Traditionally, analysis of CT chest scans was performed manually by skilled radiologist who recognized areas of abnormal airway properties in consecutive slices of the examined scan. However, analyzing about 400 slices covering the

* This work was supported by "Fundación Reina Mercedes" from "Hospital Universitario Virgen del Rocío" (Sevilla), and "Consejería de Salud de la Junta de Andalucía". Carlos S. Mendoza was supported by a doctoral scholarship financed by Universidad de Sevilla.

chest area is very tedious and too cumbersome for everyday clinical use. Moreover, manual analysis performed by the radiologist was only qualitative estimation of airway abnormalities without accurate quantification of pathological changes.

Even though airway tree abnormalities can be detected based on 2D slices, the ability to extract a full 3D model of the airway tree from a 3D image has several key advantages. For example, slice based measurements can be inaccurate if the airway is not perpendicular to the slice. Also, information available from the slices is deprived of useful context, making it harder for a radiologist to keep track of the generation number of an airway, the structure of nearby airways or the overall shape of a segment [3].

There has been a number of efforts to try to delineate the airway tree in chest CT scans. Airway tree segmentation is a complex task, mainly due to inhomogeneous grey level of the voxels located inside the bronchial lumen, artefacts caused by blood vessels adjacent to airway walls and changes of intensity levels along airway walls.

Many airway segmentation methods use region-growing algorithms, which attempt to separate air and soft tissue voxels using an HU (Hounsfield Unit) threshold [4–8]. Region growing is fast and assumes no prior knowledge of the shape or size of the airways. Choosing an appropriate global HU threshold is difficult, however, as the lungs are filled with air and misclassifying a single wall voxel can allow the segmentation to leak into the lung parenchyma.

Other methods make use of grey-level morphological operators [9–12], or front wave propagation schemes [13–15], to impose structural properties derived from a priori anatomical knowledge.

Region growing is the preferred method for initializing several of the aforementioned algorithms, being a fast and intuitive technique [3,10,11]. Some recent segmentation techniques perform region growing stages on images derived from the intensity of the scan, like the morphological gradient [16], or the posterior of a classification stage [17]. Thus, fully automatic region growing approaches become mandatory, specially those that can provide as much information as possible for the following algorithmic stage, with very low leakeage.

Few fully automatic region growing methods have proven successfull in these tasks over a significant variety of imaging devices. One of the most common techniques for threshold determination was proposed by Mori et al. [6]. Their approach, widely known as explosion-controlled region growing, tries to determine optimal thresholds by detection of sudden volume increase in the segmented region. The difficulty lies in defining how much volume increase is considered explosion, as compared to normal region growth. In our work, we also propose a self-assessed region growing algorithm. Our assessment strategy is founded on a previous normalization stage, and makes use of an iteratively computed contrast measure. In our method there is no need to define any confidence margins, as the segmented region is determined precisely by the evolution of the proposed contrast measure.

Apart from manual trial-and-error adaptive threshold selection [18], some self-assessed adaptive region growing strategies have been proposed in the past, in a variety of application contexts.

In their work [19], Hojjatoleslami and Kittler proposed a method based on finding the global maxima for two different contrast measures which they computed iteratively, as intensity-decreasing pixels were added to the segmented region. The success of the assessment was founded on the assumption that maximal contrast occurred on region boundaries, which is a reformulation of approaches assuming that the variation of the gray values within regions is smaller than across regions, an inherent assumption in all region growing techniques [20]. Unfortunately, the exhaustivity of their per-pixel approach entailed very low computational efficiency, aggravated by their multiple complex peripheral measures. Revol-Muller et al. [21] used morphological measures to assess the multiplier of the adaptive range in region growing. Instead of computing their assessment function for every pixel addition to the region, they sampled the function for an evenly-spaced set of values.

In our method we propose an assessment function based on a simple measure of the evolving contrast for the region growing sequence. To make this approach computationally feasible in 3D, we produce only evenly-spaced samples of this function, along the values of the assessed parameter defined on the normalized dynamic range of the image. This sampling strategy dramatically reduces computational complexity while preserving most critical values.

Globally, our goal is to provide fast automatic segmentation based only on the location of the upper part of the trachea. Our automatic segmentation algorithm extracts a large fraction of the visible airways with few false positive branches. For some applications, however, even this accuracy rate can be insufficient. Our method can then be considered as an initialization stage for other refining strategies, in a wide variety of imaging devices, and with no significant leakage, as suggested by our experimental results.

2 Method

2.1 Airway Lumen Intensity Model

Since our goal is providing a mechanism for airway segmentation with minimal user intervention, we have established a model that takes into account their intensity distribution in CT images.

We model then our object of interest as a connected region whose pixel intensities are sampled from a Gaussian distribution with unknown mean and standard deviation. We know that our region of interest is surrounded by other tissues derived from other, sometimes adjacent, intensity distributions. Although common in the literature, this assumption for the intensities is rarely met in practice, in the sense that the intensity distributions of tissues are only approximately Gaussian, as can be inferred from direct observation of histograms. Besides, partial overlap between adjacent distributions often occurs. For dealing with these inconveniences we propose the use of an assessment function, that

is to be evaluated along a sequence of region growing stages (region growing sequence).

2.2 Segmentation Algorithm

Normalization and Denoising Since our method was conceived for images from a wide range of scanners and acquisition protocols, we have developed a normalizing stage that accounts for such variability. As we will introduce later on, for the self-assessed region growing stage of the algorithm, we require the input intensity dynamic range to be normalized with respect to some parameter estimates of the objective intensity distribution.

In the following equations in which we describe the normalization process, N is a cubic neighborhood of radius R around the seed, $\boldsymbol{x}$ is a voxel position, $f(\boldsymbol{x})$ is the intensity for voxel at $\boldsymbol{x}$, $\bar{f}_N$ is the mean intensity estimate in N and $|N|$ is the cardinality of N. Moreover, σ_{f_N} is the estimated standard deviation for intensities in N, K is a constant parameter, and $f(\boldsymbol{x})$, $f'(\boldsymbol{x})$ are the input and output intensities for the non-linear mapping described below.

$$\bar{f}_N = \frac{1}{|N|} \sum_{\boldsymbol{x}_k \in N} f\left(\boldsymbol{x}_k\right) \quad , \tag{1}$$

$$\sigma_{f_N} = \sqrt{\frac{1}{|N|} \sum_{\boldsymbol{x}_k \in N} \left(f\left(\boldsymbol{x}_k\right) - \bar{f}_N\right)^2} \ , \tag{2}$$

$$f'(\boldsymbol{x}) = \left(1 + \exp\left(-\frac{f(\boldsymbol{x}) - \bar{f}_N}{\left(\frac{K\sigma_{f_N}}{3}\right)}\right)\right)^{-1} \quad . \tag{3}$$

In a first step we proceed by Gaussian distribution maximum-likelihood (ML) estimation of the mean and standard deviation as in (1-2), and then perform a non-linear normalization using a sigmoidal transfer function centered on the estimated mean as in (3). The width of the sigmoidal window extends $K\sigma$ around the center $\bar{f}_N$ of the mapping. For $K = 3$ the width of the window would be enough to map 99.7% of the samples, of a Gaussian distribution with similar mean and standard deviation. Greater values of K ensures robust mapping for the estimated distribution (that of the tissues of interest). The sigmoidal mapping has been chosen because of its smoothness, and its ability to focus the output dynamic range on a given input intensity range of interest.

Finally, we perform non-linear denoising using an in-slice bidimensional median filter with kernel radius Γ. Other denoising schemes would be valid, always keeping in mind that stronger smoothing usually involves loss of smaller airways.

Self-Assessed Region Growing Departing from a normalized and filtered version of the image under study, whose intensities lie in the range $[0, 1]$, we apply our self-assessed contrast-maximizing algorithm. The initial region that

needs to be provided (the upper trachea) can be obtained by means of manual seeding, or by automatic detection using one of the many methods available in the literature [22, 23]. For this initial implementation we decided to simply provide the seeds manually, as described below.

Considering an initial region R_0 defined by several seeds along the upper trachea, we provide the analytical description of the i-th iteration of the algorithm:

1. Update multiplier $k_i = k_0 + i\Delta k$
2. Compute, in last iteration grown region R_{i-1}, ML estimates for the mean (available from last iteration) and standard deviation $(\bar{f'}_{R_{i-1}}, \sigma_{f'_{R_{i-1}}})$
3. For every candidate voxel $\boldsymbol{x}_{c_{i-1}}$ being 26-connected to R_{i-1}, $\boldsymbol{x}_{c_{i-1}} \in R_i$ if

$$f'\left(\boldsymbol{x}_{c_{i-1}}\right) \in \left[\bar{f'}_{R_{i-1}} \pm k_i \sigma_{f'_{R_{i-1}}}\right] \tag{4}$$

4. Compute the assessment function $O_i\left(\bar{f'}_{R_i}, \bar{f'}_{P_i}\right)$ using the intensity average $\bar{f'}_{R_i}$ in R_i and the intensity average $\bar{f'}_{P_i}$ in the external perimeter P_i of R_i according to (1) and the following eqs.:

$$P_i = \{x_{c_i}\} \cap {R_i}^C \ , \tag{5}$$

$$O_i\left(\bar{f'}_{R_i}, \bar{f'}_{P_i}\right) = \left|\frac{\bar{f'}_{P_i} - \bar{f'}_{R_i}}{\bar{f'}_{P_i} + \bar{f'}_{R_i}}\right| \tag{6}$$

5. If O_{i-1} was a local maximum, when compared to O_{i-2} and O_i (only when $i \geq 2$), then the algorithm stops and the output is R_{i-1}. Otherwise another iteration takes place

Of all aforementioned parameters only k_0 and Δk are critical for the performance of the algorithm. k_0 affects computational efficiency requiring a greater number of iterations before a local maximum of $O\left(\bar{f}_{R_i}, \bar{f}_{P_i}\right)$ is actually found. Therefore, its fine tuning for a specific scanner, could save some computational time. From observation of the region growing sequence, we conclude that these first iterations are typically very fast, so the improvement is frequently negligible. In what concerns Δk, the choice must guarantee that the assessment function is being sampled adequately in order to detect its local variations. Since the estimates for the mean and standard deviation are continually updated as the region grows, the estimates become increasingly close to the theoretical values. We argue that setting Δk below one tenth of 3 (which is the theoretical value multiplying the standard deviation of a Gaussian distribution for 99.7% of its samples to be included in a range of that width around the mean) is enough for the segmentation process to be able not to miss the available local maxima of the assessment function. This claim is supported by our experimental results.

3 Results

We have implemented our algorithm using open source medical image processing libraries, more precisely the Insight Toolkit [24] for algorithm development, and

the command line executable module infrastructure provided by 3DSlicer for fast prototyping, calibration, evaluation, and manual segmentation on real images for further validation [25]. The algorithm that we will validate, uses the following parameter values: $R = 2, K = 12, \Gamma = 1, k_0 = 1$ and $\Delta k = 0.1$. The values for these parameters were determined from other non-thoracic CT images, and are intended to suit any imaging conditions. No thoracic CT scans were use for the tuning of these parameters, as the algorithm was initially conceived for general purpose segmentation. The algorithm was implemented and executed in a 2 GHz Intel Core 2 Duo Windows PC with 2 GB RAM, and the average running time was 129 ± 27 s.

For generation of the presented airway segmentations, the initial region was provided using three manually selected seeds along the upper trachea. The segmentation process took between 1 and 2 minutes for each dataset, including the reading/writing of the images.

The evaluation of this algorithm has taken place in comparison with other 14 algorithms using a set of 40 CT scans, with different acquisition parameters. First 20 datasets were used for training (unnecessary in our approach), and last 20 for testing. The produced segmentations were centrally evaluated by a team of trained observers. The objective was to compare performance. For this purpose, a ground truth was constructed from all available segmentations from the different algorithms, and all results were subsequently evaluated with respect to this ground truth.

Evaluation of each individual segmentation is performed in the following steps:

1. Airway segments or branches were extracted from submitted airway tree segmentations using a fast marching based algorithm [14].
2. Airway segments were evaluated visually on a set of extracted slices from both a reoriented view and a reformatted view with straightened airway centerlines.
3. Each segment was scored as "correct" or "wrong", by at least two observers. The criterion used is whether the extracted airway segment indeed belongs to the airway tree; the exact airway shape and dimensions are not taken into account.

The ground truth is then defined as the union of all valid airway segments from all submitted segmentations.

The following measurements are computed and used for comparing the submitted results:

1. Branch count: The number of branches that were detected correctly. A branch is considered detected as long as the length of the centerlines is more than 1 mm.
2. Branch detected: The fraction of branches that were detected, with respect to the branches present in the ground truth.
3. Tree length: The sum of the length of the centerlines of all correctly detected branches.

Table 1. Evaluation measures for the twenty cases in the test set.

	Branch count	Branch detected (%)	Tree length (cm)	Tree length detected (%)	Leakage count	Leakage volume (mm^3)	False positive rate (%)
CASE21	89	44.7	46.0	41.6	0	0.0	0.00
CASE22	54	14.0	37.0	11.2	8	2085.7	30.67
CASE23	33	11.6	27.3	10.5	0	0.0	0.00
CASE24	49	26.3	43.2	26.6	0	0.0	0.00
CASE25	83	35.5	63.5	25.2	0	0.0	0.00
CASE26	22	27.5	15.8	24.0	0	0.0	0.00
CASE27	35	34.7	26.0	32.1	0	0.0	0.00
CASE28	56	45.5	40.5	37.0	0	0.0	0.00
CASE29	74	40.2	44.4	32.2	0	0.0	0.00
CASE30	44	22.6	30.2	19.8	0	0.0	0.00
CASE31	77	36.0	53.7	30.6	3	31.1	0.35
CASE32	80	34.3	62.4	28.6	2	314.3	2.81
CASE33	83	49.4	56.9	38.7	0	0.0	0.00
CASE34	266	58.1	189.8	53.1	2	39.3	0.18
CASE35	112	32.6	78.4	25.3	0	0.0	0.00
CASE36	59	16.2	54.1	13.1	0	0.0	0.00
CASE37	46	24.9	39.2	22.0	0	0.0	0.00
CASE38	35	35.7	26.9	40.5	0	0.0	0.00
CASE39	93	17.9	73.8	18.0	4	65.8	0.95
CASE40	40	10.3	30.9	8.0	0	0.0	0.00
Mean	71.5	30.9	52.0	26.9	0.9	126.8	1.75
Std. dev.	51.7	13.1	36.5	11.8	2.0	466.5	6.84
Min	22	10.3	15.8	8.0	0	0.0	0.00
1st quartile	40	17.9	30.2	18.0	0	0.0	0.00
Median	58	33.4	43.8	25.9	0	0.0	0.00
3rd quartile	89	44.7	63.5	38.7	2	39.3	0.35
Max	266	58.1	189.8	53.1	8	2085.7	30.67

4. Tree length detected: The fraction of tree length in the ground truth that was detected correctly.
5. Leakage count: The number of unconnected groups of "correct" regions that were neighboring with a "wrong" region. Indicates how easy/difficult it is to manually separate leakages from the correctly detected branches.
6. Leakage volume: The volume of regions that were wrongly detected.
7. False positive rate: The fraction of the volume of regions that were detected wrongly over the volume of all detected regions.

Trachea was excluded from the branch length and branch count related measurements. For the leakage based measures, both trachea and main bronchi were excluded. In Table 1 we provide all the computed experimental measures. We also present 3D renderings for the 20 datasets used for validation, in Fig. 1. These results prove that our algorithm is able to produce reasonably complete segmen-

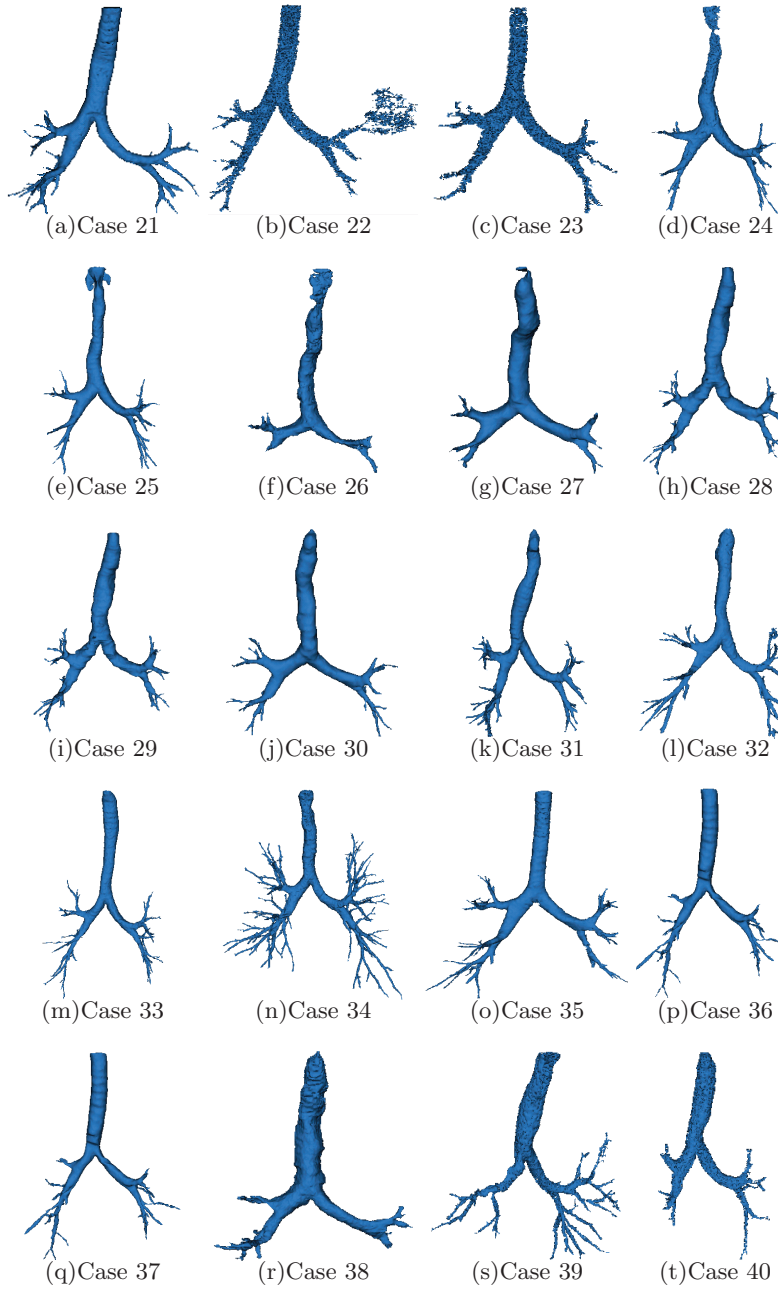


Fig. 1. (a)-(t). Segmented airways for test cases 21 through 40

tations, with very limited leakage. In most of the available test sets, the maximal contrast condition has produced fully automatic segmentations which compare fairly with those obtained using more sophisticated methods. The results were obtained with the same parameters for all different acquisition varieties, in a reduced time frame.

4 Concluding Remarks and Future Work

Our experimental results suggest that our modified region growing strategy, when used as an initialization stage, might benefit many airway segmentation algorithms currently available. Also, in those techniques which exploit region growing approaches applied over images obtained through processing of the image intensities, the use of our maximal-constrast stopping condition could be useful for further automatization.

Our immediate future line of work will include further study of our novel region-growing stopping criterion. We intend to develop an exhaustive comparison between previously available explosion detection and our newer approach. Further along the way, we would like to develop a hybrid method, using our contrast-assessed region growing for initialization, and adding some refining stages using morphological operations and surface interpolation. This will provide for further comparison between the different initialization strategies.

References

1. Sluimer, I., Schilham, A., Prokop, M., Ginneken, B.V.: Computer analysis of computed tomography scans of the lung: A survey. IEEE Transactions on Medical Imaging **25**(4) (2006) 385–405
2. Coxson, H.O., Rogers, R.M.: Quantitative computed tomography of chronic obstructive pulmonary disease. Academic Radiology **12**(11) (2005) 1457–1463
3. Szymczak, A., Vanderhyde, J.: Airway segmentation by topology-driven local thresholding. In: Progress in Biomedical Optics and Imaging - Proceedings of SPIE. Volume 6914. (2008)
4. Mori, K., Hasegawa, J.I., Suenaga, Y., Toriwaki, J.I.: Automated anatomical labeling of the bronchial branch and its application to the virtual bronchoscopy system. IEEE Transactions on Medical Imaging **19**(2) (2000) 103–114
5. Summers, R.M., Feng, D.H., Holland, S.M., Sneller, M.C., Shelhamer, J.H.: Virtual bronchoscopy: Segmentation method for real-time display. Radiology **200**(3) (1996) 857–862
6. Mori, K., Hasegawa, J., Toriwaki, J., Anno, H., Katada, K.: Recognition of bronchus in three-dimensional X-ray CT images with application to virtualized bronchoscopy system. Proceedings of the 13th International Conference on Pattern Recognition **3** (1996) 528–532
7. Kiraly, A.P., Higgins, W.E., Hoffman, E.A., McLennan, G., Reinhardt, J.M.: 3D human airway segmentation for virtual bronchoscopy. In: Proceedings of SPIE - The International Society for Optical Engineering. Volume 4683. (2002) 16–29

8. Singh, H., Crawford, M., Curtin, J.P., Zwiggelaar, R.: Automated 3D segmentation of the lung airway tree using gain-based region growing approach. MICCAI **2** (2004) 975–982
9. Fetita, C.I., Prteux, F., Beigelman-Aubry, C., Grenier, P.: Pulmonary airways: 3-D reconstruction from multislice CT and clinical investigation. IEEE Transactions on Medical Imaging **23**(11) (2004) 1353–1364
10. Graham, M.W., Gibbs, J.D., Higgins, W.E.: Robust system for human airway-tree segmentation. In: Progress in Biomedical Optics and Imaging - Proceedings of SPIE. Volume 6914. (2008)
11. Sonka, M., Park, W., Huffman, E.A.: Rule-based detection of intrathoracic airway trees. IEEE Transactions on Medical Imaging **15**(3) (1996) 314–326
12. Aykac, D., Huffman, E.A., McLennan, G., Reinhardt, J.M.: Segmentation and analysis of the human airway tree from three-dimensional X-ray CT images. IEEE Transactions on Medical Imaging **22**(8) (2003) 940–950
13. Tschirren, J., Huffman, E.A., McLennan, G., Sonka, M.: Intrathoracic airway trees: Segmentation and airway morphology analysis from low-dose CT scans. IEEE Transactions on Medical Imaging **24**(12) (2005) 1529–1539
14. Schlathölter, T., Lorenz, C., Carlsen, I.C., Renisch, S., Deschamps, T.: Simultaneous segmentation and tree reconstruction of the airways for virtual bronchoscopy. In: Proceedings of SPIE - The International Society for Optical Engineering. Volume 4684 I. (2002) 103–113
15. Ginneken, B.V., Baggerman, W., Rikxoort, E.M.V.: Robust segmentation and anatomical labeling of the airway tree from thoracic CT scans. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) **5241 LNCS**(PART 1) (2008) 219–226
16. Fabijanska, A.: Two-pass region growing algorithm for segmenting airway tree from MDCT chest scans. Computerized Medical Imaging and Graphics (2009) Article in Press.
17. Lo, P., Bruijne, M.D.: Voxel classification based airway tree segmentation. In: Progress in Biomedical Optics and Imaging - Proceedings of SPIE. Volume 6914. (2008)
18. Adams, R., Bischof, L.: Seeded region growing. IEEE Transactions on Pattern Analysis and Machine Intelligence **16**(6) (1994) 641–647
19. Hojjatoleslami, S.A., Kittler, J.: Region growing: A new approach. IEEE Transactions on Image Processing **7**(7) (1998) 1079–1084
20. Haralick, R.M., Shapiro, L.G.: Image segmentation techniques. Computer Vision, Graphics, & Image Processing **29**(1) (1985) 100–132
21. Revol-Muller, C., Peyrin, F., Carrillon, Y., Odet, C.: Automated 3D region growing algorithm based on an assessment function. Pattern Recognition Letters **23**(1-3) (2002) 137–150
22. Hu, S., Hoffman, E.A., Reinhardt, J.M.: Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. IEEE Transactions on Medical Imaging **20**(6) (2001) 490–498
23. Sluimer, I., Prokop, M., Ginneken, B.V.: Toward automated segmentation of the pathological lung in CT. IEEE Transactions on Medical Imaging **24**(8) (2005) 1025–1038
24. Yoo, T.S., Ackerman, M.J., Lorensen, W.E., Schroeder, W., Chalana, V., Aylward, S., Metaxas, D., Whitaker, R.: Engineering and algorithm design for an image processing API: a technical report on ITK–the Insight Toolkit. Studies in health technology and informatics **85** (2002) 586–592

25. Pieper, S., Lorensen, B., Schroeder, W., Kikinis, R.: The NA-MIC Kit: ITK, VTK, pipelines, grids and 3D Slicer as an open platform for the medical image computing community. In: 2006 3rd IEEE International Symposium on Biomedical Imaging: From Nano to Macro - Proceedings. Volume 2006. (2006) 698–701

3D segmentation of the airway tree using a morphology based method

Benjamin Irving, Paul Taylor, and Andrew Todd-Pokropek

University College London, WC1E 6BT, UK,
b.irving@ucl.ac.uk

Abstract. Segmentation of the airways is useful for the analysis of airway compression and obstruction caused by pathology. This paper outlines an automatic method for segmentation of the airway tree. This method includes algorithms to detect the trachea, segment the trachea and main bronchi by thresholding and region growing, and segment the remaining bronchi by morphological filtering and reconstruction. Morphological filtering and reconstruction are applied to all slices in the axial, sagittal and coronal planes and are used to extract the smaller airways. Bounded space dilation with a leak removal restriction is applied as a region growing method. This method was evaluated on 20 cases as part of the MICCAI pulmonary image analysis workshop. The mean number of branches detected as a percentage of possible branches was 43.5%, the mean tree length detected as a percentage of the entire tree length was 36.4% and the false positive rate – that is, the percentage of the total volume that was incorrectly segmented – was 1.27%

1 Introduction

This paper outlines a morphology based algorithm for the segmentation of the airways in three dimensions from computed tomography (CT) images. The aim was to develop a flexible method for extraction of the airways from CT images. This tool will be required to model airway changes in paediatric patients with pulmonary tuberculosis leading to compression and deformation of the airways. A morphology based airway segmentation method was chosen. This method was applied to adult patients for the MICCAI pulmonary image analysis workshop and, therefore, adult patients are the focus of this report.

This method further develops the use of morphological reconstruction to extract the airways reported previously by Aykac et al [1] and Pisupati et al [2]. Both Aykac et al and Pisupati et al perform greyscale morphological reconstruction on each CT slice to extract the regional minima and, therefore, enhance the airways (discussed in Section 2.3). The greyscale morphological reconstruction is applied to each slice using a range of marker images that have been created from greyscale closing the original image with a range of structuring elements (SE). The SEs are chosen to be a similar size to the cross section of the airways present in the slice in order to enhance them. The reconstructed image is

subtracted from the original image and a threshold is applied. 2D seeded region growing is then applied slice by slice to extract the airways from the binary volume [1, 2]. Pisupati et al [2] applied their method to CT images of canine lungs while Aykac et al [1] applied their method to CT images of human lungs.

Morphology based methods have also been applied by Preteux et al [3] and Fetita et al [4]. Fetita et al [4] use an algorithm that calculates the connection cost between points in an image – that is, the smallest threshold that will provide a binary path between two points – to enhance the airways in CT slices. Once enhanced, the images are not thresholded and 3D reconstruction is performed on the greyscale images using a local energy minimisation algorithm. This energy function takes into account the tree topology of the airways as well as the greyscale intensity. There are a number of other non-morphology based methods have been used to segment the airways.

This study uses a similar morphology based method to that of Aykac et al [1] and Pisupati et al [2]. These methods have, however, been extended with the aim of improving segmentation of smaller airways and airways parallel to the slice, as well as reducing leaks. The extensions include three dimensional morphological filtering and leak removal using 3D dilation. The following section discusses the segmentation algorithm, which includes an initialisation step, a step to extract the trachea and main bronchi, a step to extract the smaller airways using morphology, and finally a region growing and leak removal step.

2 Methods

An overview of the algorithm is shown in figure 1. Note that thresholding and seeded region growing is used to segment the trachea and left and right main bronchi. It is possible to use morphology based segmentation – the method used for the rest of the segmentation – for the larger branches as well, however, this means larger structuring elements (SE), increasing the likelihood that areas of the lung with a circular appearance will be segmented.

2.1 Initialisation

This method requires an initialisation point at the start of the trachea. To select this point automatically, the slices are orientated using the DICOM header information and the lung boundaries are roughly segmented by thresholding for air regions in the coronal plane through the centre of the patient. Air filled regions were classified as regions below -500 HU. Certain cases did not conform to the HU scale and, therefore, if the slice contained only values above -500 HU then the threshold was set to *MinVal+500*. In this algorithm 10 slices above the upper most point of the segmented lung region was chosen as the starting axial slice for the segmentation. This step is included because the large variation in the region covered by a CT scan, makes it difficult to select an appropriate starting slice without knowing the location of the lung fields. After an axial slice has been selected, the trachea is found in that slice by applying a threshold to identify air

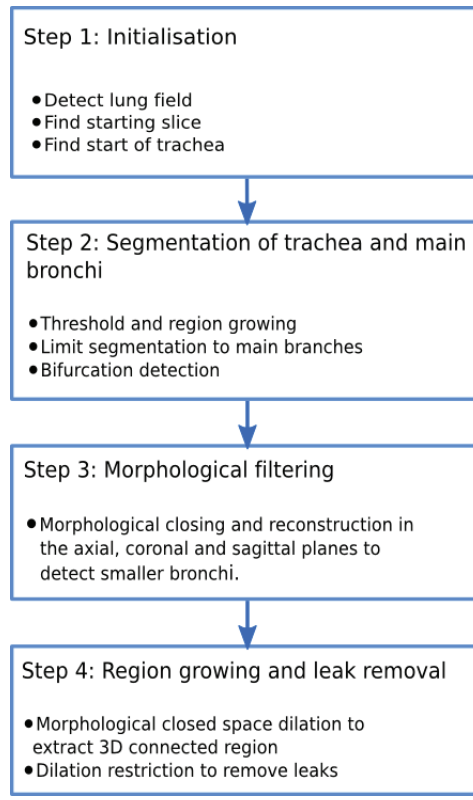


Fig. 1. Outline of algorithm to segment the airway tree

filled regions. The trachea is detected by comparing the possible regions in terms of position, size and compactness in a similar way to that of Aykac et al [1] and Mori et al [5]. The area parameter was calculated as the area of the object once the optimal area – chosen to be 400 pixels – was subtracted. The position of the object was defined as the euclidean distance from the ideal position – which was chosen to be (200, 256). Compactness (C) was calculated from perimeter (P) and area (A) as $C = \frac{P^2}{A}$. The segmentation of the starting slice is used as a seed for the rest of the segmentation.

2.2 Segmentation of Trachea and Main Bronchi

Thresholding and region growing are applied to the axial slices using the segmentation from the previous slice as seed points. This is applied progressively to each slice from the start of the trachea.

Greyscale thresholding is used only to segment larger branches, because if smaller branches are included in the segmentation, the similar greyscale intensities of the small bronchi and lungs will cause leaking into the lungs. Therefore, for each step, the newly segmented region is evaluated and steps are taken to stop the segmentation of smaller vessels. If the segmented region is more than double the area of the previous slice then a leak is considered to have occurred through smaller bronchi into the lungs. This leak is excluded by finding the centre-of-mass (CM) and the maximum radius of the segmented region in the previous slice, and iteratively reducing the radius and recalculating the CM until a specified compactness is reached; this is used to restrict the segmentation of the current slice (see figure 2).

The position of the bifurcation of the trachea and main bronchi is estimated by labelling the number of connected regions in each slice that have been seeded by one connected region in the previous slice. If two separate regions are seeded by one connected region in the previous slice then bifurcation is considered to have occurred. When the trachea bifurcates, each of the main bronchi are followed and when the main bronchi bifurcate then the procedure is stopped.

2.3 Morphological Filtering

To detect bronchi beyond the main bronchi, morphological greyscale reconstruction is applied to all CT slices to extract local minima and, therefore, enhance the airways.

Greyscale reconstruction is an extension of binary reconstruction. Binary reconstruction is the application of successive dilations within objects of a binary image as shown below [6, 7]:

$$\rho_B(X) = \lim_{n \to +\infty} \delta_B^{(n)}(X) \tag{1}$$

where $\delta_B^{(1)}(X) = (X \oplus K) \bigcap B$ and $\delta_B^{(n)}(X) = \delta_B \circ \delta_B \circ ... \delta_B(X)$ for n times, and the marker (X) is a subset of the mask (B). The marker (X) is made up of seed

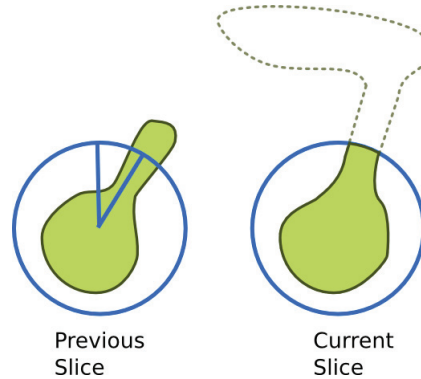


Fig. 2. Excluding leaks in trachea and main bronchi segmentation by reducing the radius and recalculating the CM iteratively until a specified compactness is reached

points – that is, where the dilation starts in the object. The mask is generally the image that is being operated on [6, 7]. K is a SE used in the dilation and $\oplus$ is the dilation operator.

This definition is expanded for greyscale reconstruction [6, 7]. A marker and mask image are used where every pixel in the marker image has a pixel intensity less than or equal to the intensity of the corresponding pixel in the mask image. The marker and mask images are then thresholded over a range of thresholds and binary reconstruction is performed on each threshold. The maximum pixel intensities of the binary reconstructed images form the greyscale reconstructed image. With a good choice of a marker image this method can be used to create a reconstructed image with the intensity peaks removed. This can then be subtracted from the original image to enhance the peaks. Figure 3 is an illustration of greyscale reconstruction in 1 dimension and shows the Mask and Marker greyscale values. The reconstructed image is represented by the shaded area.

In this study, the greyscale closing of the image of interest is used as the marker image, as shown below, in a similar way to that of Aykac et al[1] and Pisupati et al [2]:

$$X = B \bullet D = (B \oplus D) \ominus D \tag{2}$$

where X is the marker image, B is the original image (used in reconstruction as the mask image) and D is the SE. D will control the shape of the marker image and, therefore, which airways are enhanced in the reconstruction, since the difference between the original image and the reconstructed image gives

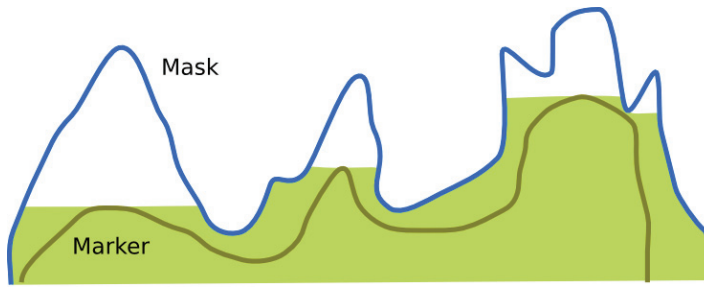


Fig. 3. Greyscale reconstruction

an airway enhanced image. Applying morphological reconstruction a number of times to a slice using different SE sizes to produce marker images will enhance airways of different cross sections [1, 2]. Therefore, morphological reconstruction was applied using marker images produced using a range of SEs. The smallest SE chosen in this study is a 4-connected binary SE. Larger SEs are created by applying successive dilations to the smallest SE i.e. $D_n = D \oplus D \oplus ... \oplus D\, n\text{times}$. The number of dilations used to create the largest SE for filtering the axial, coronal and sagittal slices were set to 12, 6 and 6 respectively. Therefore, for each slice in each orientation SEs from a 4-connected binary SE to an SE after 6 or 12 dilations are used to generate marker images.

As described earlier, reconstruction is applied to the range of marker images for each slice. Each reconstructed image is subtracted from the original image to enhance airways. These subtracted images are thresholded and the union of the thresholded images provides a binary slice with all possible airway locations. The threshold fraction used for the axial, coronal and sagittal slices were set to 0.3, 0.4 and 0.4 respectively. The threshold value is obtained from this threshold fraction in terms of the minimum and maximum greyscale values in the reconstructed slice i.e. $ThreshValue = ThreshFrac * (MaxPixelValue - MinPixelValue) + MinPixelValue$. Morphological filtering was applied slice by slice to the volume.

Aykac et al [1] and Pisupati [2] applied the filtering to each axial slice. We apply 3D filtering i.e. to each slice in the axial, coronal and sagittal plane. This is because of poor detection of branches parallel to the slice if just one direction is used. The axial plane is filtered last in order to enhance segmentation of bifurcation areas that can appear large and non circular from the axial plane if branches are parallel to the slice. Smaller branches parallel to the axial plane are, therefore, segmented first leaving more circular bifurcation areas that are detected with the axial filter (see figure 4).

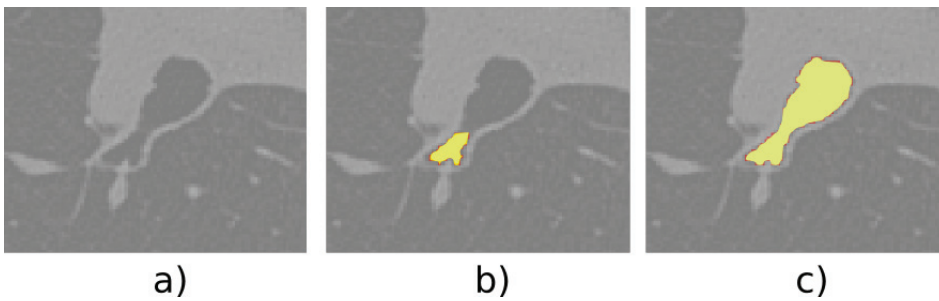


Fig. 4. Using 3D filtering to segment objects that are non-circular in the axial plane. a) Unsegmented section of the airway tree. b) Parallel vessels segmented using coronal and sagittal filtering. c) Axial filtering applied.

2.4 Region Growing and Leak Removal

Once possible regions have been detected using morphological filtering and reconstruction, bounded space dilation is applied to the binary volume as a region growing technique from the initial seed point [8].

Closed space dilation is particularly useful because airway branching can be monitored while the region growing is taking place. Closed space dilation is described as follows [8]:

$$X \oplus_B K = X_N = (X_{N-1} \oplus K) \bigcap B = ... \tag{3}$$

where $X = X_0$ is the initial seed, B is the region being segmented, K is the kernel and $\oplus$ is the morphological dilation operation [8]. Closed space dilation acts like the dilation operator except it is restricted to the shape of the object B.

This method is seeded by the initial slice in the trachea, and 3D closed space dilation is applied until the segmentation volume does not increase using a $3 \times 3 \times 3$ SE. 2D closed space dilation, which is used previously [1], is applied to each slice individually and requires a number of forward and backward passes through the whole volume to capture the 3D data. While 3D dilation allows any complexity of topology to be followed, we specify a constraint that the object remains 6-connected.

In some cases other objects that are of a similar size and shape to that of the airways are enhanced by the morphological reconstruction and remain when thresholded. If after thresholding, the object is 6-connected to the airway region it will be segmented; this causes leaks. Leaking is prevented by applying

a restriction on the dilation. The change in cross section of each branch can be monitored by calculating the area of the "growth front" of each branch, where the growth front is the area added to the connected region for each iteration of the closed space dilation [8]. The growth front of each branch is calculated by labelling in 3D each connected object of the growth front and labelling the connected growth front from the previous step. To detect leakage where the volume increase substantially, the volume of the last 3 dilations for each branch are compared to the 3 dilations before that. If the ratio of the volumes is above a specified threshold of double the size then closed space dilation along this branch is stopped.

3 Results

Figure 5 shows an example airway segmentation (visualised using ITK-SNAP). The algorithm parameters were adjusted using the 20 training images and these parameters were fixed for the 20 test images. The segmentations of the test images were submitted to EXACT09 for analysis. A ground truth was established by the organisers by taking the union of all the correct branches from all the submissions. Each submission was then compared to this ground truth. Table 1 shows the accuracy of the algorithm for the 20 test cases provided. The meaning of each heading is as follows: *branch count* is the number of branches detected correctly, *branch detected* is the proportion of branches detected compared to the ground truth, *tree length* is the sum of all correctly detected branches, *tree length detected* is the tree length compared to the ground truth, *leakage count* is the number of correct regions bordering incorrect regions, *leakage volume* is the volume of regions wrongly detected, and *false positive rate* is the fraction of wrongly detected regions out of all detected regions. This algorithm produced a fair result; the mean number of branches detected was 43.5%, the tree length detected was 36.4% and the number of false positives was 1.27%. The time taken to process a case varied according to computer specifications and the number of slices containing airways. The mean and standard deviation of the time taken to complete each of the 20 cases was 71 ± 18 minutes using a single core of a quad core 2.83GHz system. This time can be considerably decreased by converting the code from Matlab to C or another compiled language.

4 Discussion

This paper discusses a method to segment the airways using morphological closing and reconstruction developed in previous studies [1, 2]. The algorithm discussed in this paper further develops these methods to detect smaller branches and reduce leaking. This is done by applying the filtering in 3D using thresholds and structuring elements optimised for each direction, a thresholding and region growing method is applied to the larger branches to reduce the required range of structuring elements and 3D bounded space dilation along with individual branch volume change monitoring to remove additional leaks.

Table 1. Evaluation measures for the twenty cases in the test set.

	Branch count	Branch detected (%)	Tree length (cm)	Tree length detected (%)	Leakage count	Leakage volume (mm^3)	False positive rate (%)
CASE21	101	50.8	55.7	50.4	0	0.0	0.00
CASE22	76	19.6	48.4	14.7	0	0.0	0.00
CASE23	33	11.6	23.9	9.2	0	0.0	0.00
CASE24	86	46.2	63.4	39.0	0	0.0	0.00
CASE25	123	52.6	91.4	36.3	3	11.6	0.06
CASE26	55	68.8	37.8	57.5	9	16.4	0.23
CASE27	83	82.2	55.6	68.7	2	95.2	1.01
CASE28	79	64.2	58.4	53.3	7	121.3	1.34
CASE29	92	50.0	62.6	45.4	3	17.9	0.19
CASE30	94	48.2	66.6	43.6	2	48.2	0.51
CASE31	91	42.5	59.0	33.6	4	1487.4	11.71
CASE32	45	19.3	35.8	16.5	5	803.2	6.94
CASE33	85	50.6	59.6	40.5	0	0.0	0.00
CASE34	140	30.6	90.3	25.2	3	32.4	0.17
CASE35	82	23.8	53.8	17.4	0	0.0	0.00
CASE36	148	40.7	147.0	35.7	1	1.9	0.01
CASE37	65	35.1	46.9	26.4	1	94.5	0.76
CASE38	73	74.5	44.1	66.3	4	57.5	0.65
CASE39	155	29.8	111.0	27.1	1	1.4	0.01
CASE40	116	29.8	80.7	20.9	5	256.1	1.71
Mean	91.1	43.5	64.6	36.4	2.5	152.3	1.27
Std. dev.	32.3	19.1	28.2	17.1	2.6	362.7	2.91
Min	33	11.6	23.9	9.2	0	0.0	0.00
1st quartile	73	29.8	46.9	20.9	0	0.0	0.00
Median	86	44.4	58.7	36.0	2	17.1	0.18
3rd quartile	123	64.2	90.3	53.3	5	121.3	1.34
Max	155	82.2	147.0	68.7	9	1487.4	11.71

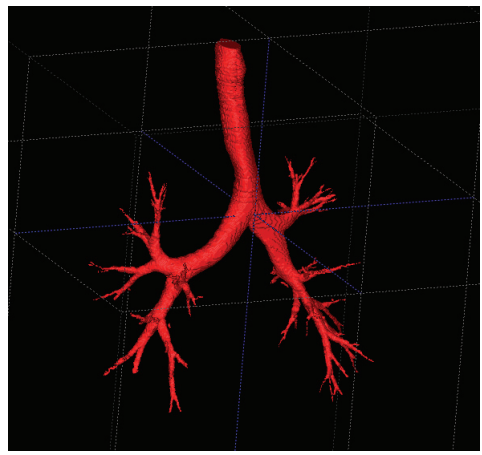


Fig. 5. Segmentation of the airways of CASE02

This method was developed for the segmentation of the airways in paediatric patients and will be used in the modelling and detection of airway deformation and stenosis. The paediatric images are of a poorer quality due to the smaller patient sizes and, therefore, fewer airway branches are visible. This algorithm could be optimised for adult airways in order to obtain better results. Branch detection will be improved by including branches that are not 6-connected to the main airway tree because of small areas of poor morphological segmentation due to airway wall discontinuities. Small discontinuities in the segmentation for CASE 22, 23, 32 and 35 removed a number of large branches when region growing was applied leading to poorer results. A more sophisticated region growing method would correct this. The benefit of this method is the low false positive rate. This is required for paediatric TB cases because exudate in the lungs increases the likelihood of segmentation leaks. Exceptions are CASE31 and CASE32 that have relatively high leakage. A further development of this method for paediatric airways includes detection of disconnected airway branches due to complete obstruction of the airways due to pathology.

References

1. Aykac, D., Hoffman, E., McLennan, G., Reinhardt, J.: Segmentation and analysis of the human airway tree from three-dimensional X-ray CT images. Medical Imaging, IEEE Transactions on **22** (2003) 940–950
2. Pisupati, C., Wolff, L., Mitzner, W., Zerhouni, E.: Segmentation of 3D pulmonary trees using mathematical morphology. Mathematical morphology and its applications to image and signal processing. Atlanta, Ga: Kluwer Academic Publishers (1996) 409–416
3. Prêteux, F., Fetita, C., Capderou, A., Grenier, P.: Modeling, segmentation, and caliber estimation of bronchi in high resolution computerized tomography. Journal of Electronic Imaging **8** (1999) 36–45
4. Fetita, C., Preteux, F., Beigelman-Aubry, C., Grenier, P.: Pulmonary Airways: 3-D Reconstruction From Multislice CT and Clinical Investigation. IEEE Transactions on Medical Imaging **23** (2004) 1353–1364
5. Mori, K., Hasegawa, J., Toriwaki, J., Anno, H., Katada, K.: Recognition of Bronchus in Three-Dimensional X-ray CT Images with Application to Virtualized Bronchoscopy System. Proceedings of ICPR **3** (1996) 528–532
6. Vincent, L.: Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. Image Processing, IEEE Transactions on **2** (1993) 176–201
7. Vincent, L.: Morphological grayscale reconstruction: definition, efficient algorithm and applications in image analysis. In: Proceedings of Computer Vision and Pattern Recognition. Volume 92. (1992) 633–635
8. Masutani, Y., Schiemann, T., Hoehne, K.: Vascular Shape Segmentation and Structure Extraction Using a Shape-Based Region-Growing Model. Lecture notes in computer science (1998) 1242–1249

A Simple Centricity-based Region Growing Algorithm for the Extraction of Airways

Rafael Wiemker, Thomas Bülow, Cristian Lorenz

Philips Research Lab Hamburg,
Röntgenstrasse 24, 22335 Hamburg
{Rafael.Wiemker, Thomas.Buelow, Cristian.Lorenz}@philips.com

Abstract. The presented algorithm was used to participate in the EXACT09 airway segmentation challenge of the Second International Workshop on Pulmonary Image Analysis (MICCAI 2009). The motivation of the presented simple algorithm mainly is to provide a benchmark what results can be achieved with very basic means in comparison to highly sophisticated algorithms. The presented algorithm uses an entirely local centricity property and an amorphous voxel-based region growing. Furthermore, it uses only raw image density value, no derivatives, or pre-processing filters. The algorithm produces quite reasonable results while being characterized by a very simple implementation, primitive data structures and quick runtime.

Keywords: EXACT09, airway extraction, bronchial tree analysis.

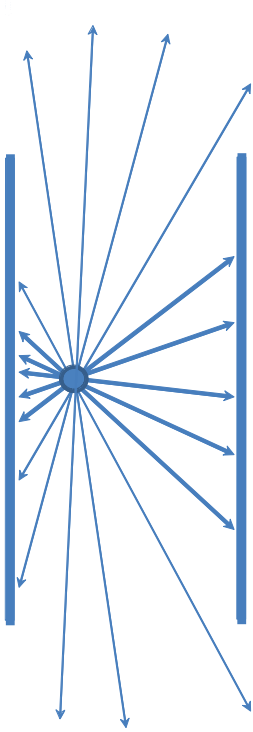
1 Motivation

The motivation of the presented simple algorithm mainly is to provide a benchmark what results can be achieved with very basic means in comparison to highly sophisticated algorithms [1-7] (using e.g. anatomical knowledge, multi-scale, multi-resolution, multi-stage, multi-rule approaches). The presented algorithm is a kind of zero-order algorithm in that it uses an entirely local centricity property and an amorphous voxel-based region growing which is not aware of segments, centerlines, branching points, directional orientation, and does not use tree-specific assumptions such as branching angles etc. It uses only raw image density value, no derivatives, or pre-processing filters.

Earlier papers with participation of the authors of this paper have been based on a front-propagation algorithm [8-14] which was use front-splitting-detection and segment branching heritage. The algorithm discussed here is not related to these earlier works.

2 Algorithmic Details

The algorithm consists only of a voxel-wise centricity measure which is used for a prioritized region growing in conjunction with two termination criteria. The centricity measure aims at quantifying how central a given voxel location is to the surrounding airway.



2.1 Local centricity measure

From a voxel position $\boldsymbol{x}$, a number of N rays is cast in a three-dimensional isotropic fashion into all directions. We used N rays sampled on the surface of a sphere according to a recursively subdivided icosahedron. The N rays are pairwise antiparallel in $N/2$ directions. The image density values are sampled in steps of dr along these rays using trilinear interpolation. Each ray is terminated if the density difference to the starting point $\boldsymbol{x}$ is higher than ΔW (assuming that this means that the airway wall has been encountered), or if becomes longer than a maximum ray length r_{max}. From each two antiparallel radii r_i and r_i' the diameter d_i is computed as $d_i = r_i + r_i'$. Out of the $N/2$ diameters d_i the $N/4$ shortest diameters are selected, i.e. all diameters below the median diameter (assuming that these diameters are approximately normal to the direction of the airway cylinder). From these selected radii the mean R and relative standard deviation σ_R / R is computed. The standard deviation of the radii becomes small or ideally vanishes if $\boldsymbol{x}$ is located centrally in the surrounding airway. Therefore we define the centricity as $c(\boldsymbol{x}) = 1 - \sigma_R / R$ (ideally 1 if $\boldsymbol{x}$ is centered in a surrounding sphere; approximating 1 for a cylinder using the median diameter selection; decreasing to < 0 for strong deviations from a cylinder).

2.2 Prioritized region growing

Starting from a seed point in the trachea (section 2.4), a three-dimensional region growing proceeds (using a 6-neighborhood) into all connected voxels below a certain density threshold D_{air} and above a certain minimum centricity value of $c(\mathbf{x}) \geq c_{min}$. The growth is prioritized by addressing the highest centricity values first.

The growth is otherwise limited only by the following rule: If the mean diameter $R(\mathbf{x})$ for a certain voxel is larger than 2 times the smallest radius encountered on the individual growth path of this voxel (along the path of predecessor voxels), then this voxel is not allowed to spawn any successor voxels.

Due to the centricity prioritization, the growth typically follows along the centerlines. As an optional post-processing step, a local dilation of the grown voxels is performed using the mean radius estimate $R(\mathbf{x})$ around each voxel.

2.3 Runtime and dual-scale-region growing

The median run-time of the algorithm on the training and test datasets was 19 seconds with a rather large standard deviation of 15 sec (on a 3 GHz single processor, using a naïve implementation of ray-casting and tri-linear interpolation). However, in a voxelwise region growing implementation (6-neighborhood) most of the runtime is actually spent to fill the relatively large volume of the trachea. Due to its simple non-delicate structure, the filling of the trachea could be done much more efficiently by other algorithms. On the other hand the paradigm of this algorithm was to keep it as simplistic as possible and not to use hybrid approaches. In a slight deviation from the simplicity-paradigm, we have modified the region-growing such that if the local radius estimation at a certain position is larger than 5 mm, then instead of a single voxel a 3×3×3 compound-voxel is grown without re-evaluation of the local centricity for each of the additionally included 26 voxels. In this way the average runtime is reduced to 5 sec without changing the results in the finer airways.

2.4 Trachea seed finding

The axial slice images are converted to binary images using a threshold of D_{air} and a two-dimensional connected component analysis is used on each slice to identify blobs with extents below 2 r_{max} . For each 2D-blob, a figure of merit is computed from its roundness and proximity to the image center. Then an iterative clustering scheme is applied which clusters blobs from adjacent slices to linear structures. These clusters are compared by virtue of mean blob-merit, blob-radius similarity, linear fit goodness, and closeness to craniocaudal orientation, and the best cluster is selected to represent a piece of the trachea, from which a central seed point is derived.

2.5 Parameters

The parameter values were not optimized on the EXACT training datasets, but adopted as sensible values from earlier datasets.

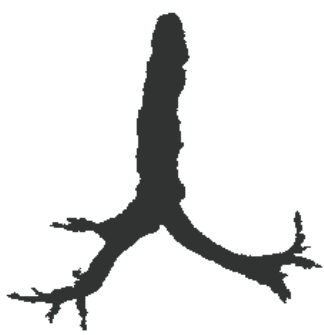
- maxAirwayDensity D_{air} = –750 HU
- maxRadiusIncrease = 2
- minimum centricity c_{min} = 0.0
- wall density difference ΔW = 200 HU
- number of rays N = 42 (antiparallel in 21 different directions)
- sampling step along rays dr = half of smallest voxel size dimension
- maximum ray length r_{max} = 20 mm (assuming that even the trachea is less than 40 mm of diameter)

3 Results

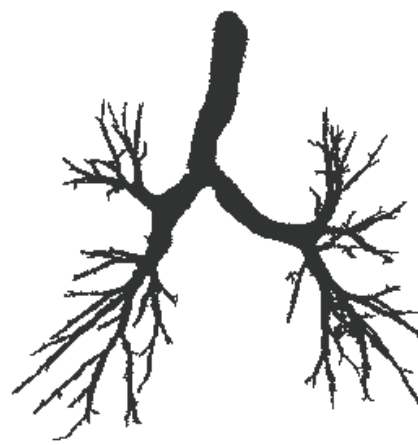
An iconing overview over training as well as testing datasets is shown below, as well as an example of worst and best case. Numerical results for all test cases computed by the EXACT09 organizers (P. Lo, M. de Bruijne, B. van Ginneken, J. Reinhardt) are given in Table 1.

"Training Data"

"Testing Data"



Example CASE38
"worst case"



Example CASE40
"best case"

Table 1: Evaluation measures for the twenty cases in the test set.

	Branch count	Branch detected (%)	Tree length (cm)	Tree length detected (%)	Leakage count	Leakage volume (mm^3)	False positive rate (%)
CASE21	118	59.3	64.7	58.5	0	0.0	0.00
CASE22	302	78.0	226.8	68.6	84	3726.2	12.12
CASE23	203	71.5	129.9	49.9	62	1184.9	5.13
CASE24	102	54.8	80.8	49.7	0	0.0	0.00
CASE25	116	49.6	86.4	34.3	0	0.0	0.00
CASE26	45	56.3	32.0	48.8	0	0.0	0.00
CASE27	41	40.6	31.2	38.6	0	0.0	0.00
CASE28	82	66.7	56.8	51.8	0	0.0	0.00
CASE29	124	67.4	79.7	57.7	7	155.7	1.58
CASE30	94	48.2	66.0	43.2	0	0.0	0.00
CASE31	128	59.8	85.5	48.7	13	742.8	4.66
CASE32	102	43.8	71.9	33.0	3	199.6	1.12
CASE33	126	75.0	95.6	65.0	7	57.1	0.69
CASE34	253	55.2	165.8	46.4	44	444.1	1.37
CASE35	143	41.6	89.4	28.9	12	82.1	0.46
CASE36	190	52.2	183.2	44.4	2	4.8	0.03
CASE37	104	56.2	71.2	40.1	2	20.5	0.15
CASE38	39	39.8	28.9	43.5	0	0.0	0.00
CASE39	237	45.6	183.7	44.9	27	690.7	3.78
CASE40	231	59.4	181.6	46.9	7	70.5	0.47
Mean	139.0	56.0	100.6	47.1	13.5	368.9	1.58
Std. dev.	73.8	11.4	57.8	10.1	23.4	854.4	2.96
Min	39	39.8	28.9	28.9	0	0.0	0.00
1st quartile	94	45.6	64.7	40.1	0	0.0	0.00
Median	121	55.7	83.1	46.6	3	38.8	0.30
3rd quartile	231	67.4	181.6	57.7	27	690.7	3.78
Max	302	78.0	226.8	68.6	84	3726.2	12.12

4 Discussion

Visual appraisal of the segmentations resulting from the algorithm clearly show a number of airway segments which can be discerned by the human eye but are not segmented by the simple algorithm. Main reasons are image noise, lack of spatial resolution (voxel sampling, slice thickness), and non-connected airway segments (caused by e.g. bronchiolitis, mucus, disease-caused alterations, anomalies, etc.).

Nevertheless, the results of the algorithm can serve as an interesting baseline for the improvements which can be achieved with more sophisticated approaches, which however, usually require many more anatomical models, rules, parameters (with its overfitting pitfalls), run-time and program code (with its maintenance costs). The charm of the algorithm stems from its simple implementation, primitive data structures and quick runtime.

References

1. S.A. Wood, E.A. Zerhouni, J.D. Hoford, E.A. Hoffman, W. Mitzner, Measurement of Three-Dimensional Lung Tree Structures by Using Computed Tomography, Journal of Applied Physiology, 79 (5), pp.1687-1697, 1995.
2. A.P. Kiraly, W. Higgins, E. Hoffman, G. McLennan, J. Reinhardt, Three-Dimensional Human Airway Segmentation Methods for Virtual Bronchoscopy, Academic Radiology, 9 (10), pp.1153-1168, 2002.
3. J.N. Kaftan, A.P. Kiraly, D.P. Naidich, C.L. Novak, A Novel Multi-Purpose Tree and Path Matching Algorithm with Application to Airway Trees, SPIE 2006 Medical Imaging Conference, SPIE vol. 6143, 2006.
4. D. Aykac, E.A. Hoffman, G.McLennan, J.M. Reinhardt, Segmentation and Analysis of the Human Airway Tree From Three-Dimensional X-ray CT Images, IEEE Trans. Medical Imaging, 22 (8), pp.940-950, 2003.
5. C. Fetita, F. Preteux, C. Beigelman-Aubry, P. Grenier, Pulmonary airways: 3D reconstruction from multi-slice CT and clinical investigation, IEEE Trans. on Medical Imaging, Vol. 23 (11), November 2004, pp. 1353-1364.
6. J. Tschirren, E.A. Hoffman, G. McLennan, M. Sonka, Segmentation and Quantitative Analysis of Intrathoracic Airway Trees from Computed Tomography Images, Proc. American Thoracic Society 2, pp.484-487, 2005.
7. J. Tschirren, E.A. Hoffman, G. McLennan, M. Sonka, Intrathoracic Airway Trees: Segmentation and Airway Morphology Analysis From Low-Dose CT Scans, IEEE Transactions on Medical Imaging, Vol. 24 (12), December 2005, pp.1529-1539.
8. Schlathölter, T., Lorenz, C., Carlsen, I.C., Renisch, S., Deschamps, T., Simultaneous segmentation and tree reconstruction of the airways for virtual bronchoscopy, Proc. SPIE 2002, vol.4684, pp.103–113.
9. R. Wiemker, T. Blaffert, T. Bülow, S. Renisch, C. Lorenz, Automated assessment of bronchial lumen, wall thickness and bronchoarterial diameter ratio of the tracheobronchial tree using high-resolution CT, Proc. Computer Assisted Radiology and Surgery, CARS 2004, pp. 967-972.
10. R. Wiemker, A. Ekin, R. Opfer, T. Bülow, P. Rogalla, Unsupervised Extraction and Quantification of the Bronchial Tree on Ultra-Low-Dose vs. Standard Dose CT, SPIE 2006 Medical Imaging, SPIE vol. 6143, 2006.
11. T. Bülow, C. Lorenz, S. Renisch: A General Framework for Tree Segmentation and Reconstruction from Medical Volume Data. MICCAI 2004, p.533-540.
12. T. Bülow, C. Lorenz, R. Wiemker, J. Honko, Point based methods for automatic bronchial tree matching and labeling, Proc. SPIE 2006 Medical Imaging Conference, San Diego, SPIE vol. 6143, 2006.
13. T. Bülow, R. Wiemker, T. Blaffert, C. Lorenz, S. Renisch, Automatic Extraction of the Pulmonary Artery Tree from Multi-Slice CT Data, Proc. SPIE 2005 Medical Imaging Conference, SPIE vol. 5746, pp.730-740, 2005.
14. Wiemker, R., Bülow T., Opfer R., Automated Hierarchical Partitioning of Anatomical Trees, Proc. SPIE 2007 Medical Imaging Conference, SPIE vol. 6511.

Fully Automated Extraction of Airways from CT Scans Based on Self-Adapting Region Growing

Oliver Weinheimer, Tobias Achenbach, and Christoph Düber

Department of Diagnostic and Interventional Radiology, Johannes Gutenberg University of Mainz, Germany
mail@oliwe.com

Abstract. The segmentation of the airway tree is an important preliminary step for many clinical applications. In this paper we present a method for fully automated extraction of airways from volumetric computed tomography (CT) images based on a self-adapting region growing process. The method consists of 3 main steps. Firstly the histogram of a dataset is analysed. Secondly the trachea is searched and segmented. And thirdly the bronchial tree is segmented by a self-adapting region growing process. The proposed method has been applied to 40 patient datasets provided by EXACT09, a comparative study of airway extraction algorithms. Former versions of our method have been used extensively in many clinical studies.

1 Introduction

Computed tomography (CT) is currently the method of choice for noninvasive and sensitive imaging of pathologic changes in the lung. Development of multidetector CT (MDCT) combines the advantages of the high-resolution CT (HRCT) and spiral CT and allows visualisation of the lungs and the bronchial tree up to the subsegmental level. A fast and reliable extraction of the airway tree is of fundamental importance for many clinical applications like a noninvasive 3D measurement and quantification of airway geometry [1, 2], computer-assisted bronchoscopy [3] or emphysema quantification [4]. Many semi-automatic and automatic methods have been presented in the past for extracting the airway tree in volumetric CT scans, e.g. [5–8]. Low dose CT scans and ultra low dose scans are increasingly utilised in lung screening studies. Lowering the radiation exposure increases the amount of noise in the CT images, hence it increases the demands on fast and reliable airway extraction methods. So far, there have been no perfect extraction techniques, however events like EXACT09 are important for comparing and improving different airway extraction methods.

2 Fully Automated Extraction of Airways from CT

Depending on the quality of the data bronchial tree extraction can be a very challenging task - especially if it is a fully automated extraction. Our proposed

method in this paper is based on the method first introduced in [9]. We improved the procedure over time while using it for many studies, see e.g. [1, 4, 10–13]. The datasets of these studies were generated with a normal clinical dose - datasets of this quality were the main field of application of former versions of our method. The datasets provided by EXACT09 range from clinical dose to ultra low dose scans, from healthy volunteers to patients with severe lung disease, and from full inspiration to full expiration. All images had a matrix size of 512×512. We additionally integrated slight modifications to our existing method so that it was applicable to the challenging datasets of EXACT09. The method consists now of 3 main steps. Firstly the histogram of a dataset is analysed. Secondly the trachea is searched and segmented. Thirdly the bronchial tree is segmented by a self-adapting region growing process.

2.1 Step 1: Histogram Analysis

The fundamental importance of calibration for a CT system is indisputable. A CT system should be well calibrated, such that air is at -1000 HU, while water is at 0 HU. For this reason the first peak in the histogram of the first 10 upper slices of a dataset is determined and assigned to the variable T_{Air}. The peak value should be -1000 HU. After temporarily applying a Gaussian filter to theses slices and recalculating the histogram, the first peak is assigned to the variable $T_{\mathrm{Air}}^{'}$. The Gaussian filter should not change the position of the first peak. We distinguish between two cases. Firstly, if $T_{\mathrm{Air}} \neq T_{\mathrm{Air}}^{'}$, see e.g. Fig. 1(a)(b), the 3×3 Gaussian mask $\frac{1}{4}[1\,2\,1] * \frac{1}{4}[1\,2\,1]^T$ is applied to each slice of the whole dataset. The inequality is caused in this case by a CT system where the range of the CT numbers is limited by -1024 HU and a hard reconstruction kernel was used, hence it is a sign of noise in the images. Secondly, if $T_{\mathrm{Air}} \neq -1000\,\mathrm{HU}$ (see e.g. Fig. 1(c)(d)), the CT system is not well calibrated to air. The air calibration error will be dealt with in step 3.

2.2 Step 2: Searching for the Trachea

The trachea search is realised on the upper slices of a data set. A body detection is performed, so that the search area for the trachea can be limited to the body region. Then a circular region with voxel values < -500 HU (dark region) is searched on the axial slices inside the detected body. The 2D region must be greater than $5^2 \times \pi\,\mathrm{mm}^2$ and smaller than $15^2 \times \pi\,\mathrm{mm}^2$. The centre of gravity of the region is calculated and mapped on the succeeding slice. The mapped point should be part of a similar dark region, additionally the top of one lung is searched on this slice. If all conditions are fulfilled, a trachea landmark is found. If no Gaussian filter was applied to the dataset in step 1, the noise is quantified in the trachea region found and if necessary the 3×3 Gaussian mask $\frac{1}{4}[1\,2\,1] * \frac{1}{4}[1\,2\,1]^T$ is applied to each slice of the whole dataset. The trachea is then segmented with a 2D region growing with threshold value -500 HU, always mapping the centre of gravity of a marked region to the succeeding slice

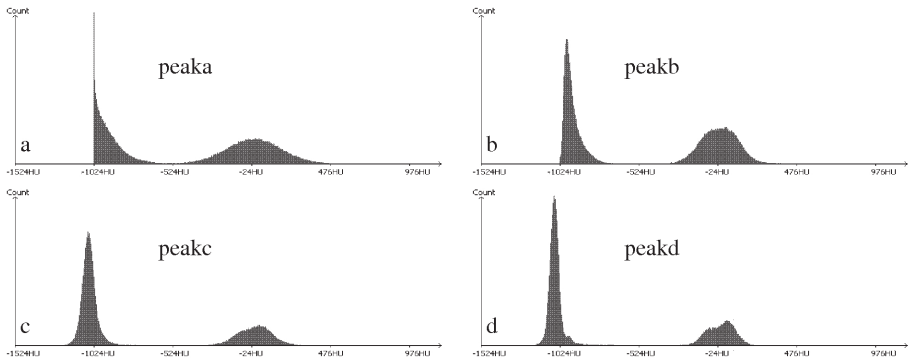


Fig. 1. (a) Histogram of CASE21 (Siemens Sensation 64, Kernel B50f, Pixel Spacing: 0.60 mm, Slice Thickness: 0.60 mm, Spacing Between Slices: 0.60 mm, Exposure: 100 mAs, 120 kVp). Values < -1024 HU are mapped on -1024 HU. Useful information is lost. (b) Histogram of CASE21 after applying a Gaussian filter. T'_{Air} is greater than -1000 HU because of the lost information described in (a). (c) Histogramm of CASE24 (Toshiba Aquilion, Kernel FC12, Pixel Spacing: 0.65 mm, Slice Thickness: 1.00 mm, Spacing Between Slices: 0.8 mm, Exposure: 5 mAs, 120 kVp). (d) Histogram of CASE24 after applying a Gaussian filter. The Gaussian filter did not change the position of the first peak. This dataset is not well calibrated, because $T'_{\text{Air}} = T_{\text{Air}} = -1055$. This indicates a calibration error of 55 HU.

in basal direction as a new seed point. The procedure stops if the carina (main bifurcation of the trachea) is reached.

2.3 Step 3: Self-Adapting Region Growing

Step 2 supplies a starting voxel within the trachea for an iterative, self-adapting and region growing based 3D bronchial tree tracer. This starting voxel is marked as "bronchusL" (L for large). All connected voxels are written in a queue. 3D region growing with a N_{26} neighbourhood system is started for all voxels added to the queue, starting with the first one. We define $T_{\text{Lumen}_{i,i=0}} = -950 + (T_{\text{Air}} + 1000)$ and $T_{\text{Wall}_{i,i=0}} = T_{\text{Lumen}} + 175$ for the first iteration. A voxel is marked as "bronchusL" if the mean value in N_7 neighbourhood $< T_{\text{Lumen}_i}$, the maximal value in $N_{27} < T_{\text{Wall}_i}$ and the voxel is connected to another voxel marked as "bronchusL". The conditions are selected so restrictively that leaking out of the segmentation into the lung parenchyma is almost impossible. If the conditions are not fulfilled, it is examined whether a voxel is in a smaller bronchus. The rationale for the 2nd evaluation is the following: a bronchus will be cut either by a axial, coronal or sagittal plane in a circular to elliptical way. If a voxel lies within a bronchus, then it is surrounded by bronchial wall in one of the planes in all directions. The algorithm does not examine all directions, but in each plane 8 rays as direction of detection are cast outwards. On each ray the maximal positive gradient and the maximal HU value are determined. Then it is determined if a voxel is air and if it is surrounded by airway wall in one plane in all

directions. The evaluation is based on fuzzy logic rules, which takes into account the average HU value, the maximal positive gradient and the maximal HU value on the rays. Again the conditions are selected so restrictively that "leaking out" into the lung parenchyma is almost impossible. If a voxel is identified as lumen voxel in one plane, the voxel is marked according to the plane in which it has been identified as "bronchusS" (sagittal), "bronchusC" (coronal) or "bronchusA" (axial). If more than a defined permitted number of voxels are added by a single region growing process, the marked voxels are reset to unmarked in order to avoid leaking out into lung parenchyma. If all voxels in the queue are processed, $T_{\text{Lumen}_{i+1}} = T_{\text{Lumen}_i} + \Delta T$ is set for the next iteration. All unmarked voxels, connected to the detected bronchial tree so far, are written in a queue and region growing starts again for all voxels added to the queue. The method stops if $T_{\text{Lumen}_{i+1}} > T_{\text{Lumen}_0} + \Delta T_{\max}$ or leakage occurred more than $L_{\max}$ times. We use $\Delta T_{\max} = 100$, $\Delta T = 1$ and $L_{\max} = 5$ by default.

After finalisation of step 3, holes in the segmented bronchial tree are closed.

3 Results and Discussion

The images used in this challenge were volumetric chest CT scans acquired at different sites using several different scanners, scanning protocols, and reconstruction parameters. The images were divided into two sets: a training set (CASE01-CASE20) and a testing set (CASE21-CASE40). We used the training set in order to slightly modify our existing method so that it was able to cope with the EXACT data. The datasets range from clinical dose to ultra low dose scans, from healthy volunteers to patients with severe lung disease, and from full inspiration to full expiration. Our proposed method was able to extract the bronchial tree fully automatic in all 40 datasets. We submitted our segmentation results to EXACT09. The segmentations of the 15 participating teams were centrally evaluated by a team of trained observers. For this purpose, a ground truth was constructed from all submitted segmentations and all submissions were subsequently evaluated with respect to this ground truth. For details on the set-up of the study, collection of data and evaluation of segmentations, acquisition parameters of the 20 test cases, and the seven evaluation measures computed see [14]. Table 1 documents the results achieved with our method for the 20 cases in the testing set. An average number of 130.1 branches were detected in the datasets (mean value for all teams: 124.01). The mean value for the leakage volume was 559.0 mm^3 (mean value for all teams: 700.55 mm^3). In 10 cases no leakage occurred - in 10 cases leakage occurred. Fig. 2(a) shows CASE32, where the greatest leakage volume was measured. Fig. 2(b) shows CASE22, where the greatest number of branches were determined. The average runtime on a PC (Intel Xeon CPU, 2.83 GHz, 4GB RAM) per case of the testing set was 183 s, 39 s for Step 1 and 2 and144 s for Step 3.

Our implemented leakage detection is just based on the number of added voxels by a single region growing process - this simple rule should be improved, shape features can be used for this purpose. It should be possible with an im-

Table 1. Evaluation measures for the 20 cases in the test set.

	Branch count	Branch detected (%)	Tree length (cm)	Tree length detected (%)	Leakage count	Leakage volume (mm^3)	False positive rate (%)
CASE21	103	51.8	54.7	49.5	3	651.4	6.93
CASE22	252	65.1	181.3	54.9	2	8.9	0.04
CASE23	154	54.2	100.6	38.7	0	0.0	0.00
CASE24	101	54.3	76.0	46.7	0	0.0	0.00
CASE25	134	57.3	100.7	40.0	0	0.0	0.00
CASE26	39	48.8	28.8	43.8	0	0.0	0.00
CASE27	34	33.7	25.0	30.8	0	0.0	0.00
CASE28	98	79.7	71.6	65.3	7	214.3	1.69
CASE29	111	60.3	69.8	50.6	5	131.3	1.04
CASE30	144	73.8	108.8	71.2	12	2226.0	12.37
CASE31	173	80.8	134.2	76.5	21	1956.2	7.38
CASE32	155	66.5	130.6	60.0	20	3563.2	10.79
CASE33	119	70.8	87.2	59.3	1	15.3	0.15
CASE34	251	54.8	158.6	44.3	0	0.0	0.00
CASE35	104	30.2	64.9	21.0	0	0.0	0.00
CASE36	129	35.4	135.1	32.8	0	0.0	0.00
CASE37	52	28.1	44.8	25.2	0	0.0	0.00
CASE38	35	35.7	27.7	41.7	0	0.0	0.00
CASE39	176	33.8	129.3	31.6	12	461.3	3.51
CASE40	238	61.2	187.0	48.3	29	1952.7	5.46
Mean	130.1	53.8	95.8	46.6	5.6	559.0	2.47
Std. dev.	66.1	16.6	49.3	14.9	8.7	1019.3	3.95
Min	34	28.1	25.0	21.0	0	0.0	0.00
1st quartile	98	35.4	54.7	32.8	0	0.0	0.00
Median	124	54.6	93.9	45.5	1	4.5	0.02
3rd quartile	176	70.8	135.1	60.0	12	1952.7	6.93
Max	252	80.8	187.0	76.5	29	3563.2	12.37

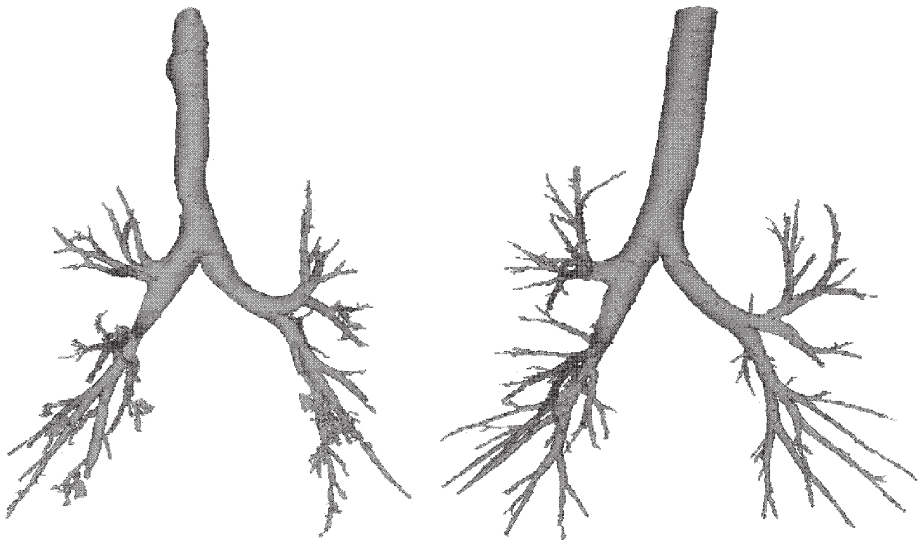


Fig. 2. Two segmented bronchial trees rendered with the marching cubes algorithm. (a) CASE32 (Philips Mx8000 IDT 16, D Kernel, Pixel Spacing: 0.78 mm, Slice Thickness: 1.0 mm, Spacing Between Slices: 1.0 mm, Exposure: 40 mAs, 140 kVp). Dataset with the greatest value for leakage volume ($3563.2\,\mathrm{mm}^3$). (b) CASE22 (Siemens Sensation 64, B50f Kernel, Pixel Spacing: 0.60 mm, Slice Thickness: 0.6 mm, Spacing Between Slices: 0.6 mm, Exposure: 100 mAs, 120 kVp). Dataset with the greatest value for branch count, 252 branches were detected, leakage volume $= 8.9\,\mathrm{mm}^3$.

proved leakage detection to increase the upper Hounsfield limit, specified by $\Delta T_{\max}$, in step 3 of the method. This should allow the segmentation of more peripheral bronchi. The rules for the detection of smaller airways in step 3 should be reworked and additionally more than the 3 main cutting planes (axial, sagittal, coronal) should be used for the decision-making process. Furthermore, 2D airway detection should be applied to a dataset and the results connected to the 3D segmentation. This can help to detect airway stenosis.

In this paper we presented a method for fully automated extraction of airways from CT Scans. Our method worked well on the challenging datasets of EXACT09, nevertheless we have gathered valuable information for our future work. Generating a common database covering a wide range of possible CT scans is an important step for improving and comparing different airway extraction methods.

References

1. Achenbach, T., Weinheimer, O., Biedermann, A., Schmitt, S., Freudenstein, D., Gouthma, E., Kunz, R.P., Buhl, R., Dueber, C., Heußel, C.P.: MDCT Assessment

of Airway Wall Thickness in COPD Patients Using a New Method: Correlations with Pulmonary Function Tests. European Radiology **18** (2008) 2731–2738
2. Saba, O.I., Hoffman, E.A., Reinhardt, J.M.: Maximizing Quantitative Accuracy of Lung Airway Lumen and Wall Measures Obtained from X-Ray CT Imaging. Journal of Applied Physiology **95** (2003) 1063–1075
3. Kiraly, A.P., Helferty, J.P., Hoffman, E.A., McLennan, G., Higgins, W.E.: Three-Dimensional Path Planning for Virtual Bronchoscopy. IEEE Trans. Med. Imaging **23**(9) (2004) 1365–1379
4. Heußel, C.P., Herth, F., Kappes, J., Hantusch, R., Hartlieb, S., Weinheimer, O., Kauczor, H.U., Eberhardt, R.: Fully-Automatic Quantitative Assessment of Emphysema in Computed Tomography – Comparison with Pulmonary Function Testing and Normal Values. European Radiology (2009)
5. van Ginneken, B., Baggerman, W., Rikxoort, E.M.: Robust Segmentation and Anatomical Labeling of the Airway Tree from Thoracic CT Scans. In: MICCAI 2008. Volume LNCS 5241.
6. Tschirren, J., Hoffman, E.A., McLennan, G., Sonka, M.: Intrathoracic Airway Trees: Segmentation and Airway Morphology Analysis from Low-Dose CT Scans. IEEE Trans. Med. Imaging **24**(12) (2005) 1529–1539
7. Fetita, C., Prêteux, F., Beigelman-Aubry, C., Grenier, P.: Pulmonary Airways: 3-D Reconstruction From Multislice CT and Clinical Investigation. IEEE Trans. Med. Imaging **23**(11) (2004) 1353–1364
8. Kiraly, A.P., Higgins, W.E., McLennan, G., Hoffman, E.A., Reinhardt, J.M.: Three-dimensional Human Airway Segmentation Methods for Clinical Virtual Bronchoscopy. Academic Radiology **9**(10) (2002) 1153–1168
9. Weinheimer, O., Achenbach, T., Buschsieweke, C., Heußel, C.P., Uthmann, T., Kauczor, H.U.: Quantification and Characterization of Pulmonary Emphysema in Multislice-CT: A Fully Automated Approach. In Petra Perner, Rüdiger W. Brause and Hermann-Georg Holzhütter, eds.: Medical Data Analysis, 4th International Symposium, ISMDA 2003, Berlin, Germany, October 9-10, 2003, Proceedings. Volume 2868 of Lecture Notes in Computer Science. Springer (2003)
10. Heußel, C.P., Kappes, J., Hantusch, R., Hartlieb, S., Weinheimer, O., Kauczor, H.U., Eberhardt, R.: Contrast Enhanced CT-Scans are not Comparable to Non-Enhanced Scans in Emphysema Quantification. European Journal of Radiology (accepted) (2009)
11. Weinheimer, O., Achenbach, T., Bletz, C., Düber, C., Kauczor, H.U., Heußel, C.P.: About Objective 3-D Analysis of Airway Geometry in Computerized Tomography. IEEE Trans. Med. Imaging **27**(1) (2008) 64–74
12. Heußel, C.P., Achenbach, T., Buschsieweke, C., Kuhnigk, J., Weinheimer, O., Hammer, G., Düber, C., Kauczor, H.U.: Quantifizierung des Lungenemphysems in der Mehrschicht-CT mittels verschiedener Softwareverfahren: Quantification of Pulmonary Emphysema in Multislice-CT Using Different Software Tools. Fortschr Röntgenstr (Röfo) **178** (2006) 987–998
13. Zaporozhan, J., Ley, S., Eberhardt, R., Weinheimer, O., Iliyushenko, S., Herth, F., Kauczor, H.U.: Paired Inspiratory/Expiratory Volumetric Thin-Slice CT Scan for Emphysema Analysis: Comparison of Different Quantitative Evaluations and Pulmonary Function Test. Chest **128** (2005) 3212–3220
14. Lo, P., van Ginneken, B., Reinhardt, J., de Bruijne, M.: Extraction of Airways from CT (EXACT'09). In: Second International Workshop on Pulmonary Image Analysis. (2009)

Multiscale Vessel-guided Airway Tree Segmentation

Pechin Lo[1], Jon Sporring[1], and Marleen de Bruijne[1,2]

[1] Image Group, Department of Computer Science, University of Copenhagen, Denmark, pechin@diku.dk,
[2] Biomedical Imaging Group Rotterdam, Departments of Radiology & Medical Informatics, Erasmus MC, Rotterdam, The Netherlands.

Abstract. This paper presents a method for airway tree segmentation that uses a combination of a trained airway appearance model, vessel and airway orientation information, and region growing. The method uses a voxel classification based appearance model, which involves the use of a classifier that is trained to differentiate between airway and non-airway voxels. Vessel and airway orientation information are used in the form of a vessel orientation similarity measure, which indicates how similar the orientation of the an airway candidate is to the orientation of the neighboring vessel. The method is evaluated within EXACT'09 on a diverse set of CT scans. Results show a favorable combination of a relatively large portion of the tree detected correctly with very few false positives.

1 Introduction

Most existing airway segmentation methods are based on region growing, with the assumption that the airway lumen has low intensity and is surrounded by higher intensity airway walls. The main problem with such an intensity based region growing algorithm is that the contrast between the airways and their surroundings is sometimes very low, due to noise or pathologies such as emphysema. Such low contrast regions often cause the region growing algorithm to leak into the surrounding lung tissue. Currently there are two approaches to address this problem: explosion control and the use of more advance image descriptors than intensity alone.

The idea of explosion control is to stop the segmentation in the low contrast regions where otherwise leakage would occur, while the segmentation continues in other regions. Strategies for explosion control generally involve heuristic rules based on geometrical properties of the regions labeled. Some examples of these geometrical properties are: volume of the regions segmented [1], radius of propagation front [2, 3], cross section area [4] and topology of thinned structure [5].

The second approach makes use of local image information to better differentiate between airways and their surroundings, for instance using pattern recognition techniques [6–8] or local tube fitting [9]. The method presented in this paper belongs to this second approach.

This paper presents an extension of our previous work [8], where we proposed to incorporate both a trained appearance model and the similarity between the orientation of an airway and its accompanying vessels. In this paper we use multi-scale Hessian eigen analysis instead of the fixed scale analysis as described in [8]. The method is evaluated within the EXACT'09 [10] airway extraction challenge on a database of 20 scans taken at different sites under a variety of different conditions. The results were manually evaluated by trained observers and compared to results submitted by other participants.

This paper is organized as follows: Section 2 explains how the airway appearance model is constructed using the training set. Section 3 presents the various steps involved in computing the vessel orientation similarity measure. The segmentation framework that combines both the airway appearance model and the vessel orientation similarity is presented in Section 4. Section 5 presents the results of the 20 cases in the EXACT'09 testing set. Finally, a discussion of the results and comparison with results submitted by other teams are presented in Section 6.

2 Classification based airway appearance model

2.1 Incomplete segmentation as a basis for training

One of the drawbacks of a classification-based appearance model is the need for training data. We have shown in [7] and [8] that incomplete but leakage free airway tree segmentations, which can be obtained relatively easily, can be used as a substitute for real ground truth segmentations as training data.

We obtain the needed manual segmentations using intensity based region growing, where both a seed point within the trachea and an intensity threshold are provided manually. The highest threshold possible without causing any leakage is selected for each training image individually. This typically results in an over-conservative segmentation that is incomplete, with many missing branches but has no leakage. As the 'background' regions directly surrounding such a conservative segmentation will always contain some airway voxels, an additional 'leaked segmentation', obtained using a slightly higher threshold, is used to take this into account. We use these leaked segmentations to prevent uncertain regions that may be either airway or background from being used in the training process. An example of a manual and a leaked segmentation is shown is Figure 1(a) and (b).

Before the extraction of training samples, we extract the lung fields, trachea and main bronchi using a thresholding and morphological smoothing based algorithm, as presented in [7, 8]. Training samples from two classes are extracted from the training data: the airway class and the non-airway class. The airway class consists of all voxels that are labeled in the manual segmentation, excluding the trachea and main bronchi. The non-airway class is limited to the area that is within the lung fields and close to the airways, which are obtained by first dilating the manual airway segmentation with a sphere of radius R_{dilate}. The

non-airway class then consists of the area within this dilated region that are not marked by the leaked segmentation.

To ensure approximately independent training samples, only a small percentage S_{sample} of the voxels belonging to the airway class are used for training. The same number of training samples is also extracted from the non-airway class. In order to prevent the large number of voxels in the larger airways from dominating the appearance model, we sample evenly along the distance from the main bronchi, measured through the segmented tree. This is done by grouping the voxels based on their distance from the main bronchi in bins with the width of each bin fixed to W, and randomly sample a total of $N = S_{sample}VW/D_{max}$ training samples from each bin, where V is the total number of airway class voxels in the manual segmentation and D_{max} is the maximum distance between a voxel in the manual segmentation to the main bronchi. To ensure that we do not sample a bin too densely, at most 50% of all voxels belonging to a bin will be included. The sampling starts at the bin furthest away from the main bronchi. If the required number of samples from a bin is larger than the number of voxels available in the bin, the remaining samples are extracted from the next available bin of shorter distance.

2.2 Airway probability

The training samples extracted are used to train a k nearest neighbor (KNN) classifier [11] to differentiate between voxels belonging to the airway and non-airway classes. An initial set of local image descriptors or features is computed from the training samples, which consists of spatial derivatives up to and including the second order, eigenvalues of the Hessian matrix (λ_1, λ_2 and λ_3, where $|\lambda_1| \geq |\lambda_2| \geq |\lambda_3|$), determinant and trace of the Hessian matrix, Frobenius norm of the Hessian matrix, and combinations of Hessian eigenvalues that measure tube, plate and blobness ($|\lambda_2/\lambda_1|$, $|\lambda_3/\lambda_1|$, $(|\lambda_1| - |\lambda_2|)/(|\lambda_1| + |\lambda_2|)$, $|\lambda_3|/\sqrt{|\lambda_1\lambda_2|}$). The partial derivatives of the image are computed at multiple scales by convolving the image with the partial derivatives of a Gaussian kernel [12], and each of the features is standardized to zero mean and unit variance.

Sequential forward feature selection [13] is used to find an optimal set of image descriptors that maximizes the area under the receiver operating characteristic (ROC) curve of the classifier. To this end, the training samples are randomly partitioned into two parts to compute the ROC curve: one third for training of the classifier and two thirds for validation. The final KNN classifier is trained using the optimal combination of features and all the training samples.

We can now estimate for each voxel in previously unseen images the posterior probability of the voxel belonging to the airway class, given a set of optimal features $\boldsymbol{x}$, using the following:

$$p(A|\boldsymbol{x}) = \frac{K_A(\boldsymbol{x})}{K} \tag{1}$$

where A is the airway class, $K_A(\boldsymbol{x})$ is the number of neighbors around $\boldsymbol{x}$ that belong to the airway class, obtained among the K nearest neighbors.

3 Obtaining vessel orientation similarity

The vessels are segmented from the lung fields, using a multi-scale Hessian eigen analysis approach. The scale for calculating the Hessian matrix is selected for each voxel independently using the scale normalized [14] Frobenius norm of the Hessian matrix:

$$\omega(\sigma_i) = \sigma_i^2 \sqrt{\lambda_1(\sigma_i)^2 + \lambda_2(\sigma_i)^2 + \lambda_3(\sigma_i)^2}$$

The local vessel scale, σ_v, is then obtained by selecting the smallest scale that corresponds to a local maximum of $\omega(\sigma_i)$ across scales. Using the Hessian eigenvalues at scale σ_v, the following criteria are used to evaluate whether a voxel belongs to a vessel or not:

Brightness:	$\lambda_1, \lambda_2 < 0$
Contrast:	$\omega \geq T_\omega$
Tubeness 1:	$(\lvert\lambda_1\rvert - \lvert\lambda_2\rvert)/(\lvert\lambda_1\rvert + \lvert\lambda_2\rvert) < T_1$
Tubeness 2:	$(\lvert\lambda_1\rvert - \lvert\lambda_3\rvert)/(\lvert\lambda_1\rvert + \lvert\lambda_3\rvert) > T_2$

A voxel is labeled as vessel when it fulfills all four criteria. The brightness criterion ensures that only voxels that are brighter than their surroundings will be selected, the contrast criterion reduces the effects of noise by ensuring a certain minimum contrast between the voxel and its surroundings, and finally the two tubeness criteria require vessels to locally resemble bright, solid cylinders. Segmentation using the vessel criteria often results in additional small, isolated regions due to noise. A connected component analysis using a 6-connected neighborhood scheme is employed to remove regions that are smaller than V_{min} voxels. Finally, the vessel centerlines are obtained using the 3D thinning algorithm presented in [15].

The vessel orientation at the centerline voxels is obtained as the eigenvector corresponding to λ_3 computed at the vessel scale σ_v. This measure is less sensitive to noise and inaccuracies in the vessel segmentation than the orientation obtained directly from the centerline itself. The orientation of an airway is extracted the same way as the orientation of a vessel, through multi-scale Hessian eigen analysis. The Hessian matrix is constructed using the airway probability image, generated using (1) where the airways resemble solid bright tube structures. Given θ as the angle between the local tube orientation at an airway candidate voxel and the orientation measured at the centerline of a vessel nearest to it, the vessel orientation similarity is defined as $s = |cos(\theta)|$. When the two orientations are similar then $s \simeq 1$, and when the orientations are perpendicular then $s \simeq 0$.

4 Segmentation framework

The airway tree segmentation is obtained using a 3D region growing algorithm, with a decision function that combines both the airway appearance model of Section 2 and the vessel orientation similarity of Section 3. An initial airway

segmentation described in Section 4.1 is used as seeds for the region growing algorithm. Figure 1(f) shows a block diagram of the proposed segmentation framework.

4.1 Initialization

The segmentation process is initialized with a coarse segmentation of the first four airway generations obtained using intensity based region growing. First, starting from the trachea and main bronchi as obtained in Section 2.1, all connecting voxels with intensity lower than a threshold T_{airway} are added. This is followed by a closing operation with a sphere mask of 3 voxels in radius. Finally, an algorithm [2] that is capable of tracking generations via bifurcation detection is applied to the smoothed segmentation, and only branches up to the fourth generation are retained. The threshold T_{airway} is dynamically determined by searching from -1000HU, with an increment of 5 HU, until the resulting initial segmentation, excluding trachea and main bronchi, is larger than 1000 voxels.

4.2 Airway segmentation

The initial segmentation obtained previously is used as seed points in a region growing process to extract the remainder of the airway tree, using the airway probability and vessel orientation similarity measures. The vessel similarity is used as a means to relax the requirements on airway probability in regions with a local tube orientation that is similar to the orientation of nearby vessels. We achieve this by using the following decision function to decide on whether to accept an airway candidate voxel $\boldsymbol{x}$ or not

$$D(p(A|\boldsymbol{x}), s) = \begin{cases} 1, & p(A|\boldsymbol{x}) \geq T_u \\ 1, & T_u > p(A|\boldsymbol{x}) \geq T_l \text{ and } s \geq T_s \\ 0, & \text{otherwise,} \end{cases} \tag{2}$$

where $p(A|\boldsymbol{x})$ is the airway probability computed from (1), s is the vessel orientation similarity of the candidate voxel, T_u, T_l and T_s are the upper probability threshold, lower probability threshold, and vessel similarity threshold respectively. The voxel $\boldsymbol{x}$ is labeled as an airway when $D(p(A|\boldsymbol{x}), s) = 1$.

4.3 Parameter settings

KNN classification was performed using the ANN library for approximate nearest neighbor searching [16]. A $K = 21$ was used, and the approximation error *eps* was set to zero to turn off the approximation part of the algorithm. For the extraction of training samples, as described in Section 2.1, the dilation radius R_{dilate} was set to 5 mm, sampling percentage S_{sample} was set to 0.05, and the bin width W was set to 3 voxels. A total of 7 scales, distributed exponentially between 0.5 mm and 3.5 mm, were used to compute the features, as well as for

the multi-scale Hessian eigen analsysis in Section 3. A contrast threshold T_ω of 100, and tubeness measure thresholds T_1 and T_2 of 0.5 were used for the vessel criteria. Detected vessel regions smaller than $V_{min} = 20$ voxels were discarded. All these parameter values are the same as those reported in [8].

The thresholds for the decision function (2) was hand tuned based on the training set, where the airway probability images were generated with KNN classifiers that were constructed in a leave-one-out manner. Our experiments with the training set showed that $T_u = 0.86$, $T_l = 0.62$ and $T_s = 0.8$ gives good results, where a significant amount of new branches not in the training data were found without any significant leakages.

5 Experiments and results

Among the 20 cases (CASE01-CASE20) in the training set from EXACT'09, manual segmentations for training were successfully obtained from all cases except one (CASE06), where leakage was observed even when using the lowest possible threshold. Therefore only 19 cases from the training set were used to train the KNN based appearance model.

The training process of our method took approximately 13 hours on a single CPU of an Intel Xeon X5355 processor (2.66 GHz), with the feature selection process occupying around 11 hours. The average computation time to segment an image in the test case was approximately 1 hour and 30 minutes. Most of the computation time was spent generating the airway probability image, which took an average of 50 minutes. The remaining computation time was mainly spent on generating the Gaussian blurred images and performing the multi-scale Hessian eigen analysis.

Table 1 shows the evaluation results of our method on the 20 EXACT'09 test images. Surface renderings of the best case and worst case according to the detected tree length are shown in Figure 1(c) and (d).

6 Discussions and conclusion

At the expense of a relatively long computation time and laborious training procedure, we obtain a favorable combination of a relatively large portion of the tree detected correctly with very few false positives. Table 1 shows that the proposed method is able to extract at least 50% of the total tree length for 70% of the cases, with a false positive rate of less than 1% for all cases. Although many branches were either extracted only partly or missed completely, it should be noted that no method was able to extract more than, on average, 77% of tree length or branches in the ground truth. Among the 15 methods that were compared in EXACT'09, 7 methods resulted in both a lower tree length and a higher false positive rate. Compared to the remaining 7 methods, our results stand out mainly by the small amount of leakage; in 11 cases there were no false positives at all, and in the remaining cases both leakage volume and false positive rate were small. Figure 1(c) shows few clear leakages even in the case

Table 1. Evaluation measures for the twenty cases in the test set.

	Branch count	Branch detected (%)	Tree length (cm)	Tree length detected (%)	Leakage count	Leakage volume (mm^3)	False positive rate (%)
CASE21	114	57.3	64.2	58.1	0	0.0	0.00
CASE22	276	71.3	227.1	68.7	1	0.2	<0.01
CASE23	186	65.5	137.9	53.0	1	15.4	0.12
CASE24	128	68.8	106.0	65.1	0	0.0	0.00
CASE25	152	65.0	116.9	46.4	0	0.0	0.00
CASE26	48	60.0	32.9	50.0	0	0.0	0.00
CASE27	49	48.5	36.9	45.6	0	0.0	0.00
CASE28	77	62.6	57.3	52.2	0	0.0	0.00
CASE29	117	63.6	81.3	58.9	1	10.0	0.12
CASE30	140	71.8	109.5	71.7	3	26.2	0.32
CASE31	159	74.3	117.8	67.1	3	39.1	0.31
CASE32	151	64.8	120.9	55.5	2	7.3	0.05
CASE33	108	64.3	81.3	55.2	0	0.0	0.00
CASE34	301	65.7	213.1	59.6	3	14.0	0.07
CASE35	136	39.5	95.2	30.8	0	0.0	0.00
CASE36	187	51.4	185.2	44.9	0	0.0	0.00
CASE37	57	30.8	46.4	26.1	0	0.0	0.00
CASE38	36	36.7	27.8	41.8	0	0.0	0.00
CASE39	253	48.7	195.1	47.7	6	37.0	0.30
CASE40	333	85.6	315.5	81.5	17	214.8	0.90
Mean	150.4	59.8	118.4	54.0	1.9	18.2	0.11
Std. dev.	85.2	13.6	75.4	13.4	3.9	48.0	0.22
Min	36	30.8	27.8	26.1	0	0.0	0.00
1st quartile	77	48.7	57.3	45.6	0	0.0	0.00
Median	138	63.9	107.7	54.1	0	0.0	0.00
3rd quartile	253	71.3	195.1	67.1	3	26.2	0.30
Max	333	85.6	315.5	81.5	17	214.8	0.90

with the largest number of detected leaks (CASE40). Our method is also the only method with an average false positive rate below 1% (0.11%) that is still able to achieve an average detected tree length of higher than 50% (54%), as shown in Figure 1(g). Among all methods with an average false positive rate below 5%, the method achieving the highest tree length, which is a semi-automated method, detected only 5% more of the total tree length at a false positive rate of 1.19%.

The presented method was originally developed for segmenting the airway tree in more standardized, low-dose cancer screening scans obtained at a single site [17, 8]. While the main approach and parameter settings used for the current work are the same as in [8], the implementation differs from this previous work in three ways. Firstly, the Hessian eigen analysis to determine vessel and airway orientations, performed at a single resolution level in [8], was replaced by a

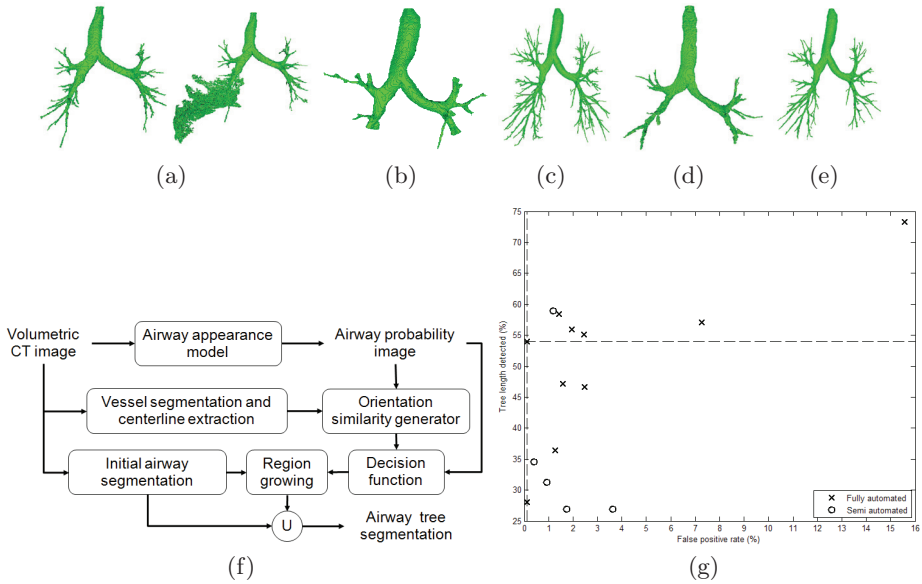


Fig. 1. Surface renderings of (a) a manual (left) and leaked (right) segmentation used for training, (b) initial segmentation, test results with (c) largest (CASE40) and (d) smallest (CASE37) percentage of tree detected. (e) Surface renderings of segmentation from CASE40 obtained using screening study scans from [8] as training. (f) Block diagram of the segmentation framework. (g) A scatter plot of average tree length detected versus average false positive rate of all participating teams in EXACT'09, with the proposed method at the intersection of the dashed lines.

multi-resolution analysis. This modification had already been developed for the screening study and improved results considerably for that data.

To cope with the much more diverse data of EXACT'09, the method was further modified in two ways from our experiments on the training images. The first modification is that a more complete segmentation is used to initialize the classification-based region growing, instead of using only the trachea and main bronchi as reported in [8]. The reason for this is because we found that otherwise the segmentation was sometimes already terminated within the first four generations in noisy images. The second modification is that the thresholds for the decision function in this work were manually tuned based on the training data, while those in our earlier work were tuned automatically using an automatic leakage detection algorithm similar to [2] and [3]. In the diverse and sometimes very noisy CT scans from EXACT'09, this rule-based, automatic leakage detection algorithm turned out to be unreliable. The criterion used in selecting the thresholds was that no obvious leakage should be present in the results on the training set, which is probably one of the reasons why our results are on the conservative side compared to other methods. A different set of thresholds may lead to longer tree lengths at the expense of an increase in false positives.

In the EXACT'09 study, a training set was provided with carefully selected images that were of similar quality as the images in the test set. If the training data does not match the test data, results may deteriorate. To illustrate this, segmentation of CASE40, obtained using a Siemens Sensation 16 scanner and very sharp convolution kernel B70s, was repeated with the same setup and same parameter settings but using the training data of [8], which consisted only of scans obtained using a Philips Mx8000 IDT 16 scanner with softer kernel D, resulting in less noisy images. The result is shown in Figure 1(e). The method trained with different data has still little or no leakage, but finds fewer branches. In this case, this was mainly because the appearance model is not capable of handling the noise and often misclassifies bright noise voxels in the airway lumen as non-airway. Although the method should be trained on similar data for optimal results, the good results on the diverse set of EXACT'09 data indicate that application of this method is not limited to studies in which acquisition conditions can be standardized.

Note that, although the proposed method requires training data, in this work only very low quality segmentations were available for this purpose. The segmentations used for training consisted of on average 93 branches and had a total length of 99 cm. Application of the trained models on the test set of similar images resulted already in much more complete segmentations, with 150 branches and a total length of 118 cm. Clearly, our method can achieve better results than the training segmentations, but the lack of training examples from small branches does limit the ability of the system to extract higher generation airways. With the availability of a set of high quality segmentations for training, such as the ground truth resulting from EXACT'09, we expect to obtain even better results in the future.

In conclusion, an airway segmentation method that uses a voxel classification based appearance model and the similarity between the orientation of an airway and its neighboring vessel is presented. Compared to the results from other algorithms submitted to EXACT'09, our method is especially effective in avoiding leakage, while still being able to extract a fair amount of airway branches.

Acknowledgments. This work is partly funded by the Danish Council for Strategic Research (NABIIT), the Netherlands Organization for Scientific Research (NWO), and AstraZeneca, Lund, Sweden.

References

1. Kiraly, A.P., Higgins, W.E., Hoffman, E.A., McLennan, G., Reinhardt, J.M.: 3D human airway segmentation for virtual bronchoscopy. In: SPIE Medical Imaging 2002: Physiology and Function from Multidimensional Images. Volume 4683. (April 2002) 16–29
2. Schlathölter, T., Lorenz, C., Carlsen, I.C., Renisch, S., Deschamps, T.: Simultaneous segmentation and tree reconstruction of the airways for virtual bronchoscopy. Volume 4684., SPIE (2002) 103–113

3. van Ginneken, B., Baggerman, W., van Rikxoort, E.: Robust segmentation and anatomical labeling of the airway tree from thoracic CT scans. In: Medical Image Computing and Computer-Assisted Intervention. Volume 5241 of Lecture Notes in Computer Science. (2008) 219–226
4. Kitasaka, T., Mori, K., Suenaga, Y., Hasegawa, J., Toriwaki, J.: A method for segmenting bronchial trees from 3D chest X-ray CT images. In: MICCAI (2). (2003) 603–610
5. Tschirren, J., Hoffman, E., McLennan, G., Sonka, M.: Intrathoracic airway trees: segmentation and airway morphology analysis from low-dose CT scans. Medical Imaging, IEEE Transactions on **24**(12) (Dec. 2005) 1529–1539
6. Ochs, R.A., Goldin, J.G., Abtin, F., Kim, H.J., Brown, K., Batra, P., Roback, D., McNitt-Gray, M.F., Brown, M.S.: Automated classification of lung bronchovascular anatomy in CT using AdaBoost. Medical Image Analysis **11**(3) (June 2007) 315–324
7. Lo, P., de Bruijne, M.: Voxel classification based airway tree segmentation. In: Medical Imaging 2008: Image Processing. Volume 6914., SPIE (2008) 69141K
8. Lo, P., Sporring, J., Ashraf, H., Pedersen, J., de Bruijne., M.: Vessel-guided airway segmentation based on voxel classification. In Brown, M., de Bruijne, M., van Ginneken, B., Kiraly, A., Kuhnigk, J., Lorenz, C., Mori, K., Reinhardt, J., eds.: Proc. of First International Workshop on Pulmonary Image Analysis. (2008)
9. Graham, M.W., Gibbs, J.D., Higgins, W.E.: Robust system for human airway-tree segmentation. In: Medical Imaging 2008: Image Processing. Volume 6914., SPIE (2008) 69141J
10. Lo, P., van Ginneken, B., Reinhardt, J., de Bruijne, M.: Extraction of airways from ct (exact09). In: Second International Workshop on Pulmonary Image Analysis. (2009)
11. Duda, R.O., Hart, P.E., Stork, D.G.: 4.4. In: Pattern Classification. 2 edn. Wiley-Interscience (2001) 174 – 177
12. Weickert, J., Ishikawa, S., Imiya, A.: On the history of Gaussian scale-space axiomatics. In Sporring, J., Nielsen, M., Florack, L., Johansen, P., eds.: Gaussian Scale-Space Theory. Kluwer Academic Publishers, Dordrecht, The Netherlands (1997) 45–59
13. Pudil, P., Novovičová, J., Kittler, J.: Floating search methods in feature selection. Pattern Recogn. Lett. **15**(11) (1994) 1119–1125
14. Lindeberg, T.: Feature detection with automatic scale selection. Int. J. Comput. Vision **30**(2) (1998) 79–116
15. Wang, T., Basu, A.: A note on 'A fully parallel 3D thinning algorithm and its applications'. Pattern Recognition Letters **28**(4) (March 2007) 501–506
16. Arya, S., Mount, D.M., Netanyahu, N.S., Silverman, R., Wu, A.Y.: An optimal algorithm for approximate nearest neighbor searching fixed dimensions. J. ACM **45**(6) (1998) 891–923
17. Pedersen, J., Ashraf, H., Dirksen, A., Bach, K., Hansen, H., Toennesen, P., Thorsen, H., Brodersen, J., Skov, B., Døssing, M., Mortensen, J., Richter, K., Clementsen, P., Seersholm, N.: The danish randomized lung cancer CT screening trial - overall design and results of the prevalence round. Journal of Thoracic Oncology (April 2009)

Segmentation of the Airway Tree from Chest CT using Local Volume of Interest

Jaesung Lee and Anthony P. Reeves

School of Electrical and Computer Engineering,
Cornell University, Ithaca, NY, USA

Abstract. Lung diseases such as COPD and asthma affect airway morphology. Automated segmentation is an essential first step toward the analysis of airways. We propose a fully-automated algorithm to segment the airway tree from chest CT scans. The proposed algorithm requires no manual intervention and uses a 3D region growing based method and allows for accurate detection of leakage by growing regions within a locally-defined envelope. The algorithm is run within a tree segmentation framework which breaks down the problem into the segmentations of individual branches. The method was evaluated using 20 chest CT scans from EXACT09 challenge. The provided scans were taken from the patients with various health conditions. The results show that the algorithm is able to segment the airway tree while keeping the leakage level low, as only 0.11% of the segmentation was classified as false positive and the average number of leakage was less than 1 per patient.

1 Introduction

The physical appearance of human airway tree is affected by lung diseases such as COPD and asthma. Technological advances in volumetric chest CT provide the opportunity to study how airways are affected by a certain disease or respond to a therapeutic treatment. Segmentation of the airway tree structure from a chest CT scan is an essential first step to further studies on airway morphology.

There has been a considerable amount of interest in segmenting airway tree structure from chest CT. The vast majority of the segmentation methods are based on 3D region growing technique. [1–6] This is due to the high contrast between airway lumen with low intensity and airway wall with high intensity. A seed point is identified inside the top-most airway segment (trachea), and the rest of the airway structure is derived from the seed point. A major problem with the growing-based approach is a leakage into the lung parenchyma region which has similar intensity as airway lumen. The airway segmentation is performed as a first step in analysis of the airway segments. The segmentation algorithm was optimized to avoid any leakage since each leakage may cause false measurements.

Although the basic principles of most growing-based algorithms are similar, the algorithms use different approaches to prevent a leak. The algorithm presented in this paper uses two levels of preventive measures to avoid a leak. First, 3D growing itself is carried out conservatively by using strict criteria for growing

a voxel. Second, leak that could not be prevented using conservative growing is subsequently detected locally and eliminated.

Since the different papers evaluate the algorithms using different datasets and evaluation metrics, it is difficult to compare different algorithms. The data used in this work is from EXACT09 (http://image.diku.dk/exact/), a segmentation challenge with the goal of comparing the results of various algorithms for extracting the airway tree from chest CT scans using a common dataset and performance evaluation method. [7]

This paper presents the details of our algorithm, provides a description of the data, and the results. In the results section, we report the performance of the presented algorithm with the evaluation metrics used in the challenge.

2 Method

The proposed algorithm has three main stages. First, a seed point is automatically identified from a given CT scan. Then, starting with the trachea using the seed point, individual branches are grown within localized cylindrical volumes. Finally, any detected leaks are removed from the segmentation.

2.1 Pre-processing

The entire CT scan is pre-filtered with 3x3 median filter to reduce noise. Kiraly et al. [5] suggested that the median filter increased the robustness of the segmentation algorithm. However, such filtering may reduce the sensitivity to identify small airway segments.

2.2 Tree Segmentation Framework

The segmentation algorithm was implemented within a framework that discovers the tree structure of the airways for subsequent segment analysis. Within this framework, each individual branch in the airway tree is represented as a node with the pointers to its parent and children, if any.

Each node may be in one of two states, “open” or “closed”. A node is in an open state when it is first created, meaning that the segmentation is not completed for the given branch. When the segmentation is complete (i.e. an end point or a bifurcation point is detected), the current node is closed, and if necessary, the child nodes are newly allocated with open states.

The growing process searches for any open node and performs segmentation on the corresponding branch. The program terminates when there are no more open nodes. Following airway segmentation airway analysis may proceed by measuring parameters of each detected airway segment.

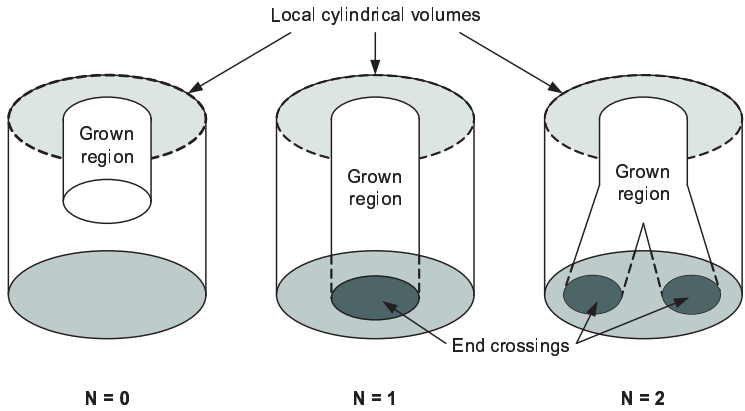


Fig. 1. Illustration of different number of end crossings (N). A view looking up at the cylinders is shown. The number of end crossings determines how the algorithm will proceed in the next iteration.

2.3 Seed Point Detection

The seed point is a point where the growing process starts and is typically located in the trachea. For this work the seed point was automatically identified in each scan from the slice 50 mm below the top-most image. The 50 mm distance was selected to be below the region of the scan where there is high noise but high enough to still be within the trachea.

The search space for a seed point was limited to a rectangular region centered on the slice of interest. The dimension of the rectangle is set to half of the scan size in x and y dimensions. Within the search space, the image is thresholded at -750 HU, and the largest connected component is found. The seed point is estimated by computing the center of mass of this component.

2.4 Local Growing

Each airway branch is grown iteratively by advancing a volume of interest. This idea is similar to the methods proposed in [3, 4]. A volume of interest is initially formed by placing a cylinder toward a given direction. The very first volume of interest is for growing the trachea and is placed on the seed point in a downward direction (d = <0,0,1>).

Each voxel is added to the segmentation if all of the following conditions are met:

1. It is connected to the currently grown voxel using 6-connectivity.
2. It has an intensity below a threshold.
3. At least half of its 26 neighbors have intensity below a threshold.

The third criterion is a modification to a typical region growing method and serves as the first level of leakage prevention.

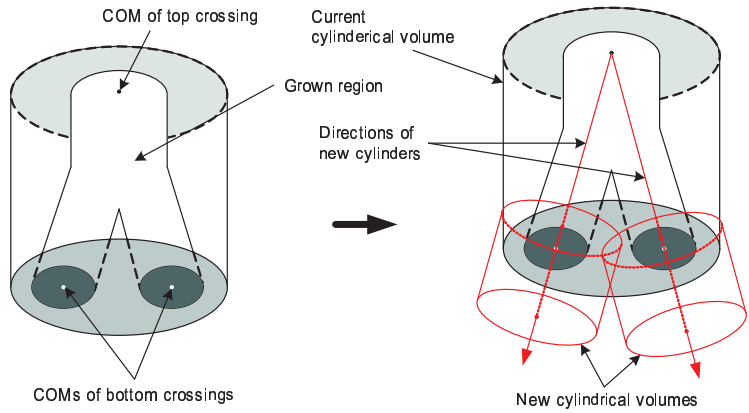


Fig. 2. Illustration of new cylindrical volume construction (N=2). A view looking up at the cylinders is shown. The left figure shows the current cylindrical volume, and the right figure shows both current and new cylindrical volume. The directions of the new cylindrical volumes are determined by connecting the center of masses of crossings.

Once a region is grown within current cylindrical volume, a new volume of interest is established based on how the grown volume crosses the current cylindrical volume. The number of crossings toward the end of the cylinder (N) determines how the growing proceeds. There are three possible categories for values of N as shown in Figure 1:

1. N = 0: The current segment has ended, and no children branches will be created.
2. N = 1: The current branch has not yet been fully grown.
3. N > 1: The current segment has ended, and children branches will be initiated.

When N is greater than 0, one or more new cylindrical volumes need to be constructed. Four parameters are needed to construct a cylinder: a start point, a direction, a diameter, and a length.

Figure 2 shows an illustration of how new cylinders are created for the case of N = 2. The direction of the newly constructed cylinder is determined using start and end crossing regions. A vector connecting the centers of masses of the start crossing and the end crossing defines the direction of a new cylinder. The starting point of a new cylinder is offset from the end crossing's center of mass in the opposite direction. This offset is introduced in order to eliminate a gap between two subsequent cylindrical volumes. The offset value is set to 10% of the cylinder's height. The diameter and length parameters were empirically determined using the training data as described in Section 2.6.

2.5 Leak Detection and Threshold Adjustment

Although growing step itself has some degree of leakage prevention, leaks may still occur. Growing airway branches iteratively allows for local detection of a leakage. Using a localized growing algorithm, a leak may be detected more accurately, and it is possible to locate where the leak has occurred.

Two criteria were used for detecting a leak in the current cylindrical volume:

1. If any of the crossing regions has a surface area greater than the α times entire cylinder's surface area. ($0.0 \leq \alpha \leq 1.0$)
2. If the grown volume is increased by β times when compared to the grown region in the previous iteration.

The probability of a leak has a direct relationship to the threshold used for growing. A high threshold makes the airway detection more sensitive but increases the chance of a leak. The optimal threshold value is determined adaptively for each case. The algorithm initially uses the threshold value of -950 HU. The threshold is incremented by 5 HU until a leak is detected. A leak detected in a cylindrical volume is removed from the segmentation output by unsetting all grown voxels within the cylinder.

2.6 Parameter Optimization

Four parameters were optimized with the training data set:

1. diameter of the cylindrical volume
2. length of the cylindrical volume
3. α - leak detection parameter (defined in Section 2.5)
4. β - leak detection parameter (defined in Section 2.5)

The only requirement for cylinder's diameter was that it must be greater than the diameter of airway branch of interest. Based on the training data, the cylinder diameter was set to 40 mm for the very first branch (the trachea) and 30 mm for all other branches. Setting the diameter to any greater value should not affect the outcome of the segmentation.

The length of the cylinder determines how much to advance in each iteration of growing. A short length means that the growing is constrained to smaller local space and is preferred since it would allow for localized detection of leakage. However, the length should be proportional to the diameter of the airway for robust placement of new cylindrical volumes. For this work, cylinder height was varied depending on branch generation. The empirically determined cylinder length based on the segmentation outcome for the training data were 20 mm for the trachea (1st generation), 17.5 mm for the main bronchi (2nd generation), 15 mm for the 3rd generation bronchi, and 10 mm for the branches with higher generations.

The values of α and β have effect on the performance of the leak detection. Smaller α and β values would mean that the leak will be detected with higher sensitivity. Based on close observation of the leakages in the training data, α and β were set to 0.5 and 4.0, respectively.

2.7 Post-processing

Once the growing process is complete (i.e. there are no open nodes), the outcome is further processed to obtain the final segmentation. The main purpose of this step is to correct any artifacts and voids that may exist due to noise present in the CT scan. First, a 3D morphological closing operation is performed on the grown binary image with a 3x3x3 spherical kernel. Then, any voids enclosed within the segmentation are filled.

3 Data

The dataset for the experiment included two sets of 20 chest CT scans provided for EXACT09 challenge. The scans were acquired at different sites using different scanners and parameters. The scans were taken from healthy volunteers as well as the patients with mild to severe lung disease and taken at various degree of inspiration and expiration. The radiation dose of the scans ranged from clinical dose to low dose.

The first 20 scans were used as a training set, and the algorithm parameters were optimized using these scans. The second set of 20 scans were used as a test set to evaluate the algorithm.

4 Results

The algorithm's performance using the evaluation metrics defined by EXACT09 challenge is shown in Table 1. The ground truth was defined as the union of all valid airway segments from all segmentations submitted to the challenge [7]. On average, the algorithm successfully segmented 81 branches (32.8% of the ground truth segmentation) with the false positive rate of 0.11%.

5 Discussion

A fully automated method for segmenting airway tree has been presented. Once an optimal threshold was found for a given case, the algorithm took less than 30 seconds to process a CT scan on a workstation with Intel Xeon 3.00 GHz CPU.

The proposed method performs segmentation in a conservative manner to prevent leakage into the lung parenchyma. When conservative growing fails to avoid a leak, the second level of leakage prevention is carried out using a local leak detector. Figure 3 shows an example of leak detection and elimination.

The algorithm was able to segment 79 airway branches for each scan on average, which corresponds to approximately the 6th to 7th branch generations. The number of successfully segmented branches corresponded to approximately one third of the number of branches in the ground truth. While the proposed algorithm may exhibit a low sensitivity, the false positive rate was very low (0.11%). It should also be noted that the number of leakages was less than 1 per scan on average.

Table 1. Evaluation measures for the twenty cases in the test set.

	Branch count	Branch detected (%)	Tree length (cm)	Tree length detected (%)	Leakage count	Leakage volume (mm^3)	False positive rate (%)
CASE21	76	38.2	43.6	39.5	0	0.0	0.00
CASE22	175	45.2	131.4	39.7	4	155.2	1.09
CASE23	152	53.5	106.8	41.0	2	26.8	0.22
CASE24	53	28.5	46.7	28.7	0	0.0	0.00
CASE25	75	32.1	58.2	23.1	0	0.0	0.00
CASE26	31	38.8	20.9	31.8	0	0.0	0.00
CASE27	33	32.7	23.9	29.5	0	0.0	0.00
CASE28	41	33.3	30.3	27.6	0	0.0	0.00
CASE29	60	32.6	36.5	26.4	0	0.0	0.00
CASE30	43	22.1	28.9	18.9	0	0.0	0.00
CASE31	59	27.6	40.6	23.1	0	0.0	0.00
CASE32	72	30.9	50.5	23.2	1	60.4	0.61
CASE33	62	36.9	44.5	30.3	0	0.0	0.00
CASE34	221	48.3	150.0	42.0	0	0.0	0.00
CASE35	99	28.8	66.5	21.5	1	42.9	0.36
CASE36	44	12.1	39.8	9.7	0	0.0	0.00
CASE37	48	25.9	40.9	23.0	0	0.0	0.00
CASE38	34	34.7	26.6	40.1	0	0.0	0.00
CASE39	110	21.2	88.1	21.5	0	0.0	0.00
CASE40	98	25.2	80.9	20.9	0	0.0	0.00
Mean	79.3	32.4	57.8	28.1	0.4	14.3	0.11
Std. dev.	51.1	9.6	36.2	8.8	1.0	37.2	0.28
Min	31	12.1	20.9	9.7	0	0.0	0.00
1st quartile	43	25.9	30.3	21.5	0	0.0	0.00
Median	61	32.3	44.1	27.0	0	0.0	0.00
3rd quartile	110	38.8	88.1	39.7	1	26.8	0.22
Max	221	53.5	150.0	42.0	4	155.2	1.09

The low leakage level was achieved by using a conservative segmentation approach. The algorithm parameters were chosen to prevent the leak as much as possible. By varying the parameters α, β, and threshold, we expect that the algorithm would achieve a higher sensitivity at the cost of increasing false positive rate (or level of leakage).

6 Conclusion

A fully-automated algorithm for segmenting airway tree from a chest CT scan has been developed and was evaluated using the data and performance metrics provided by EXACT09 challenge. The presented method segmented one third of the branches in the ground truth and exhibited an average leakage count of less than one per case.

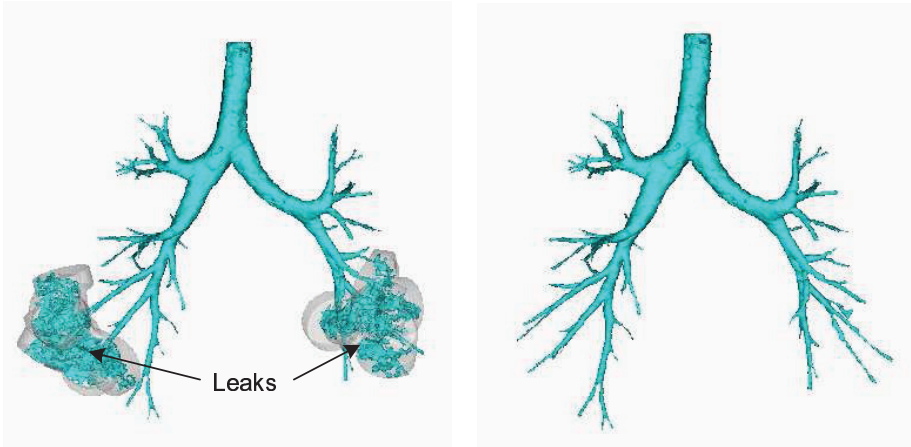


Fig. 3. An example of leak handling (CASE23). The visualization on the left shows the segmented airway with the transparent cylindrical volumes in which the leaks were detected. The right image is the visualization of the final segmentation after elimination of the detected leaks.

References

1. Graham, M., Gibbs, J., Higgins, W.: Robust system for human airway-tree segmentation. In: Proceedings of SPIE. Volume 6914. (2008) 69141J
2. van Ginneken, B., Baggerman, W., van Rikxoort, E.: Robust segmentation and anatomical labeling of the airway tree from thoracic CT scans. International Conference on Medical Image Computing and Computer-Assisted Intervention **11**(1) (2008) 219
3. Tschirren, J., Hoffman, E., McLennan, G., Sonka, M.: Intrathoracic airway trees: segmentation and airway morphology analysis from low-dose CT scans. Medical Imaging, IEEE Transactions on **24**(12) (2005) 1529–1539
4. Kitasaka, T., Mori, K., Hasegawa, J., Suenaga, Y., Toriwaki, J.: Extraction of bronchus regions from 3D chest X-ray CT images by using structural features of bronchus. In: International Congress Series. Volume 1256., Elsevier (2003) 240–245
5. Kiraly, A., Higgins, W., McLennan, G., Hoffman, E., Reinhardt, J.: Three-dimensional human airway segmentation methods for clinical virtual bronchoscopy. Academic Radiology **9(10)** (Oct 2002) 1153–68
6. Wood, S., Zerhouni, E., J.D.Hoford, Hoffman, E., Mitzner, W.: Measurement of three-dimensional lung tree structures by using computed tomography. Journal of Applied Physiology **79(5)** (1995) 1687–1697
7. EXACT09. http://image.diku.dk/exact/. Accessed July 1, 2009.

Automatic segmentation of the airway tree from thoracic CT scans using a multi-threshold approach

Eva M. van Rikxoort, Wouter Baggerman and Bram van Ginneken

Image Sciences Institute, University Medical Center Utrecht, The Netherlands

Abstract. A method for automatic extraction of the airway tree from thoracic CT scans is presented that uses adaptive thresholds while growing the airways. The method is evaluated on 20 volumetric chest CT scans provided by the Extraction of Airways from CT 2009 (EXACT09) challenge. The scans were acquired at different sites, using several different scanners, scanning protocols, and reconstruction parameters. There are scans of clinical dose, low dose, and ultra-low dose data, in inspiration and expiration, from both relatively healthy and severely ill patients. The results show that the method is able to detect a large number of airway branches at the cost of relatively high leakage volume.

1 Introduction

Multi-slice CT scanning technology has revolutionized the in vivo study of the lungs and motivates the need for pulmonary image analysis [1]. Automated extraction and labeling of the bronchial tree from thoracic CT scans is vital to accurately quantify airway morphology which is increasingly used to measure progression and response to treatment for a variety diseases. Another important application is computer-assisted bronchoscopy.

A wide variety of methods have been developed to segment the airways [2–11]. Some of these include or focus specifically on anatomical labeling of airway segments [2, 12, 13]. Most methods have been evaluated on a small number of scans. Evaluation on low-dose scans is rare ([14] is an exception), as are application to expiration scans and scans with substantial pathology.

In this work, the method as presented in [15, 16] is applied. The approach is based on the generic tree extraction framework outlined in [4, 6] and introduces several modifications and new rules for accepting segments. A key contribution is the introduction of a multi-threshold approach to increase robustness. Results are presented on the 20 test scans as provided by the Extraction of Airways from CT 2009 (EXACT09) challenge, which contain scans acquired at different sites, using several different scanners, scanning protocols, and reconstruction parameters. Scans are available from clinical dose to ultra-low dose, from full inspiration to full expiration, and from healthy subjects as well as diseased patients.

2 Method

The backbone of the algorithm is an implementation of the framework given in [4, 6]. In this section we discuss the initialization, briefly review the framework, describe the rules for accepting voxels, fronts and segments, and introduce the multi-threshold extension of the method.

Initialization. The trachea and the lungs are automatically segmented with the method described in [17]. Central dark circular regions are searched to find a start point in the trachea, followed by region growing with multiple optimal thresholds to extract the trachea and the lungs. The lung segmentation is used to infer the scan orientation. From the trachea segmentation a seed point is determined in the axial slice that contains the center of gravity of the structure. Only growing in the basal direction is allowed.

Tree segmentation framework. The segments of the bronchial tree are obtained by wavefront propagation. The initial seed point provides the first front. At every iteration, all unprocessed voxels connected to the front that satisfy the *voxel criteria* form the new wavefront. The segment is allowed to keep growing when the front meets the *wavefront criteria*. If the new front consists of multiple parts, a segment is complete and accepted if it complies with the *segment criteria*. To avoid spurious front splittings due to noise, a large 80-connectivity value is used to detect them. New fronts are pushed on a stack and the next front from the stack is propagated. The algorithm terminates when the stack is empty. While the fronts propagate, the centerline or skeleton of the tree and the local segment diameter are computed and this information is used in several of the acceptance criteria. An important difference with [4, 6] is that we use region growing to obtain the new front. To avoid diamond- or cuboid-shaped fronts, growing is restricted to within a sphere from the last calculated center point with a diameter slightly larger than the last calculated segment diameter.

Rules for accepting voxels, wavefronts and segments. Voxels are accepted when their density (in Hounsfield Units or HU) is below a threshold t, or (to be less sensitive to noise) the HU value in a $3 \times 3 \times 3$ neighborhood around the voxel is $< t$. For every new front, three checks are applied to the segment grown so far, and if they are violated the entire segment is removed. First, the segment's current radius must be smaller than 1.5 times the minimum radius found in any parent segment. This ensures that diameters of bronchi diminish. When leaking occurs, this rule is typically violated. Second, a front is not allowed to touch any other segment (segments are grown in a breadth first fashion). Third, the length of the segment should not be more than 5 times its radius. This ensures that partly grown segments are accepted before a leakage occurs that could discard a large part of an airway. A completed segment is only accepted if it meets two more requirements: The angle it makes with its parent should be $< 100^\circ$ and the average ratio of radii of two consecutive fronts should not exceed 1.1. The latter check ensures that the segment is not widening, which typically indicates leakage.

Post processing. After the bronchial tree has been extracted, several post processing steps are performed. First all minor trailing segments (i.e. segments with-

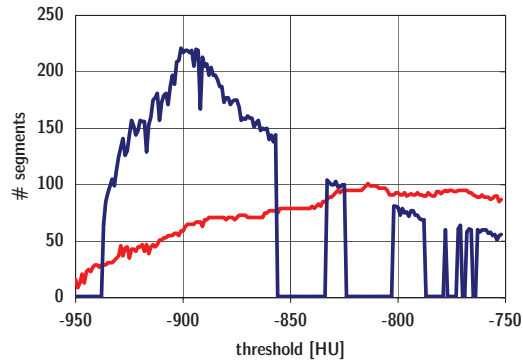


Fig. 1. The total number of segments found in the airway tree as a function of the threshold t for voxel acceptance for two scans. The blue line demonstrates that the results can be sensitive to small changes in t. For both cases, the multi-threshold method found many more segments: 289 and 181 for the blue and red line, respectively.

out descendants) are removed. Segments are considered minor if their length is smaller than 3 mm and their volume is below 25 mm^3. Next, the tree structure is scanned for segments that have exactly one descendant, and these segments are merged. Finally, holes in the segments, primarily caused by noise, are filled.

Adaptive acceptance rules. We have observed that with these rules for accepting wavefronts and segments, leakage into the parenchyma is minimal, and thus all segments found are true airways. Not all airways are found, however, and in fact it is possible that large parts of the airway tree are missed. Quite often, small changes in the value for the voxel acceptance threshold t, have a profound effect on the number and total length of detected airways. Note that it is not the case that higher values for t automatically lead to more voxels considered airways. Surely, more voxels are accepted when computing a new front when t is increased, but these fronts or these segments may subsequently be rejected by the front and segment acceptance rules. The algorithm thus manifests a complex interplay between the rules at various levels. The effect is illustrated in Fig. 1. This figure also shows that the optimal value for t varies from scan to scan.

To overcome these limitations and obtain a more robust segmentation that includes as many peripheral airways as possible, the process is made adaptive. Every segment is first grown with a high threshold $t = -750$. If rejected, the segment is regrown with a lower threshold $t + k\Delta t$ with $k = 1, \ldots, 18$ and $\Delta t = -10$. This is referred to as the multi-threshold method. It renders the extraction process adaptive: at every position in the scan the maximum number of airway voxels are selected while the front and segment rules still ensure that no leaking can occur.

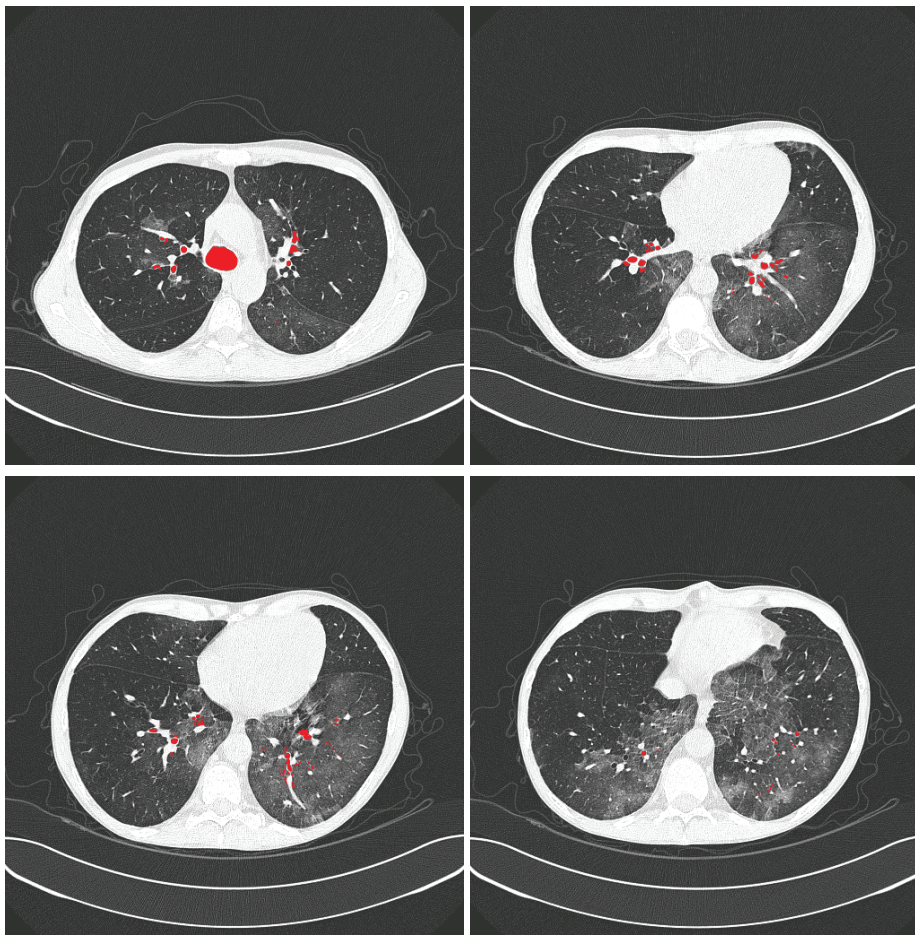


Fig. 2. Four axial slices of the case for which the most leakage occurred (case 40).

3 Experimental results

The method was applied to all 20 scans in the test set of the EXACT09 segmentation challenge. Segmentation of the bronchial tree took 10 seconds per scan on a single core PC. The results were submitted to the EXACT09 website, where the evaluation was performed. Table 1 shows the results of our method. It can be seen from Table 1 that our method is able to detect 161.4 correct airway branches per scan on average, however, on average we detect only 67.2% of all correct branches available in the ground truth. The main problem of the method seems to be leakage, with an average of 1873.4 mm^3 and a standard deviation of 2630.0 mm^3. However, there are two scans for which no leakage occurred. Figures 2, 3, 4, and 5 show several slices of the best and worst results obtained by our method.

Table 1. Evaluation measures for the twenty cases in the test set.

	Branch count	Branch detected (%)	Tree length (cm)	Tree length detected (%)	Leakage count	Leakage volume (mm^3)	False positive rate (%)
CASE21	179	89.9	88.7	80.2	15	331.1	3.57
CASE22	199	51.4	122.5	37.1	92	1784.4	7.64
CASE23	173	60.9	107.5	41.3	107	1495.2	7.11
CASE24	134	72.0	100.5	61.8	47	2446.0	9.08
CASE25	175	74.8	141.0	56.0	46	1261.4	4.59
CASE26	55	68.8	41.2	62.6	1	1.9	0.03
CASE27	54	53.5	35.6	44.0	0	0.0	0.00
CASE28	97	78.9	69.9	63.7	5	36.5	0.43
CASE29	147	79.9	108.2	78.3	37	685.6	5.72
CASE30	150	76.9	113.8	74.5	11	127.6	1.21
CASE31	165	77.1	117.4	66.9	47	2519.5	12.90
CASE32	130	55.8	93.3	42.8	42	2381.0	11.13
CASE33	135	80.4	106.8	72.6	46	529.0	5.66
CASE34	313	68.3	193.1	54.0	148	5140.4	14.22
CASE35	282	82.0	217.1	70.2	67	1775.7	7.31
CASE36	155	42.6	146.8	35.6	5	30.3	0.25
CASE37	151	81.6	108.8	61.2	21	611.8	3.66
CASE38	44	44.9	31.1	46.8	0	0.0	0.00
CASE39	322	61.9	239.2	58.5	83	5630.3	19.18
CASE40	168	43.2	126.2	32.6	62	10680.8	31.66
Mean	161.4	67.2	115.4	57.0	44.1	1873.4	7.27
Std. dev.	75.8	14.5	54.4	14.7	40.2	2630.0	7.83
Min	44	42.6	31.1	32.6	0	0.0	0.00
1st quartile	130	53.5	88.7	42.8	5	36.5	0.43
Median	153	70.4	108.5	59.8	44	973.5	5.69
3rd quartile	199	80.4	146.8	72.6	83	2519.5	12.90
Max	322	89.9	239.2	80.2	148	10680.8	31.66

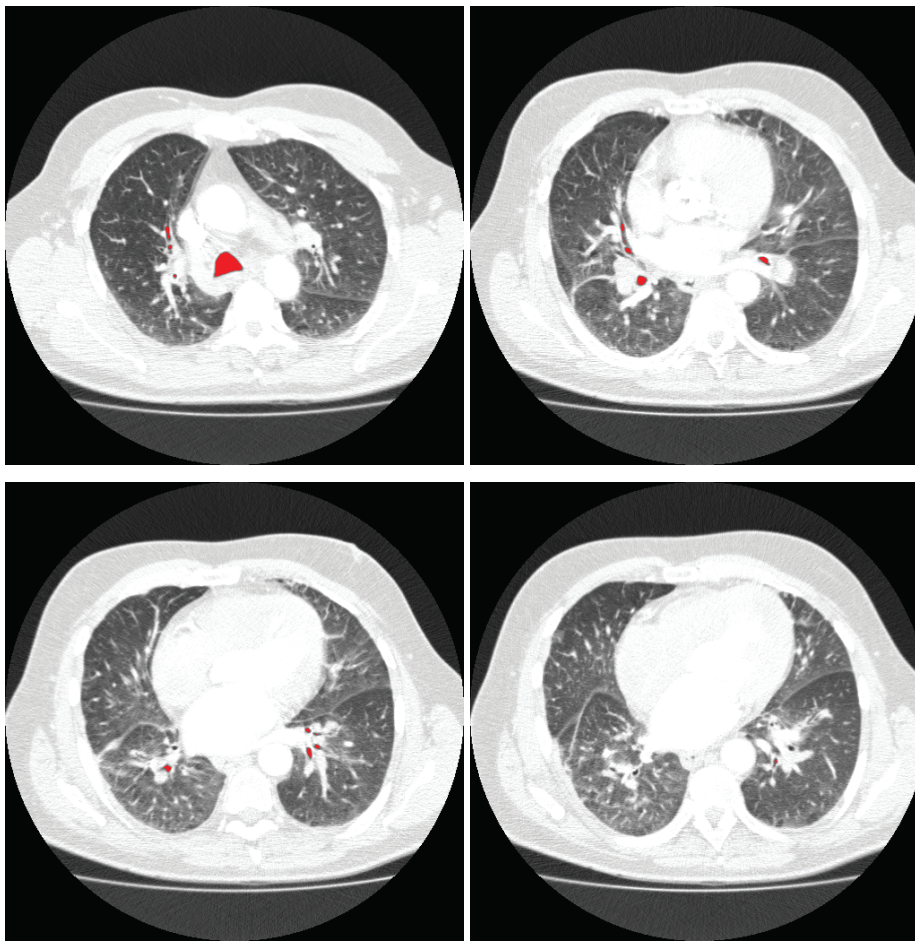


Fig. 3. Four axial slices of a case for which no leakage occurred (case 27).

4 Discussion & conclusion

In this paper, a previously proposed method for airway segmentation [15, 16] was applied to the data from the EXACT09 challenge without changes of the method or parameters. The multi-threshold method effectively uses different HU thresholds for different parts of the bronchial tree. This strategy was designed to avoid leakage, and thus early termination of the growth process. However, as can be seen in Table 1 and Figure 2, this is not always effective. The method is able to find a large percentage of correct airway branches, but at the cost of a high average leakage volume. The method takes around 10 seconds on a single-core PC.

Two main problems of the methods become apparent from the results presented in Table 1. First, the method leaks into the lung parenchyma for a large

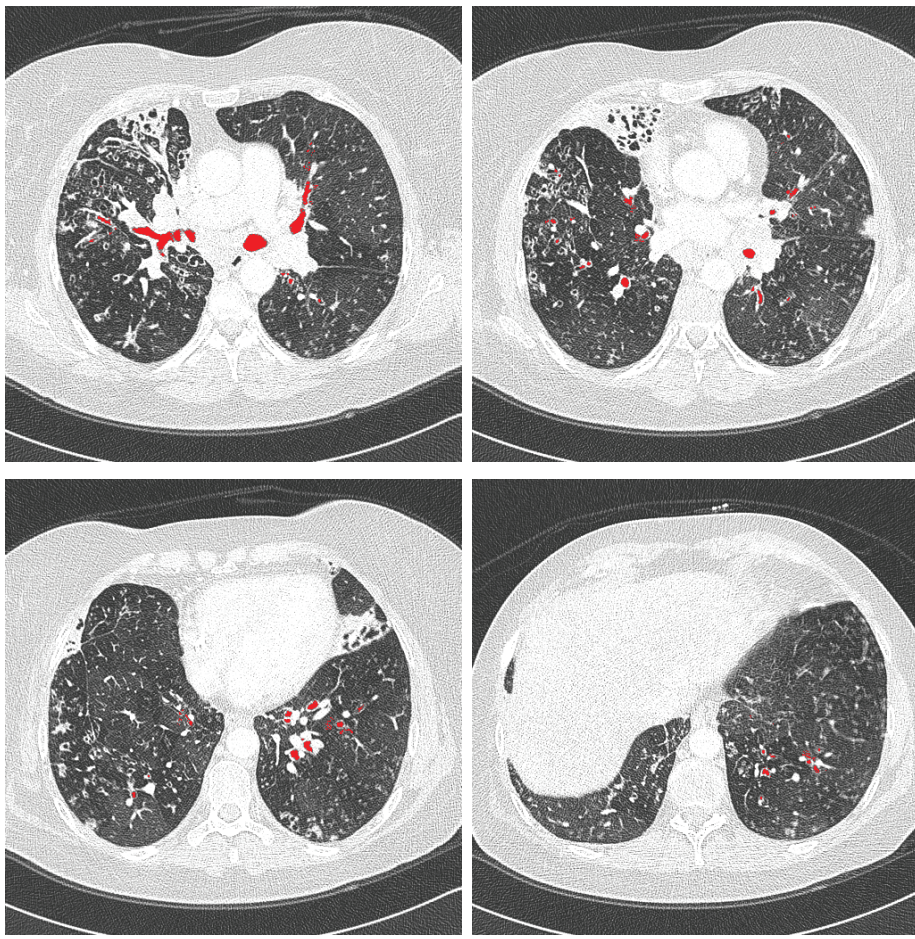


Fig. 4. Four axial slices for the case for which the most branches were detected (case 39).

percentage of scans, and second, the method misses approximately 33% of all true airway branches. To avoid leakage into the parenchyma, a leakage detection algorithm could be build into the system. In addition, leakage might be prevented by using more complex rules for accepting voxels, that are not only based on Hounsfield Units but also include statistical information about the voxel and its surrounding.

In all scans, there were (small) peripheral bronchi were present that were not extracted. A specific search for more distal airways and a mechanism to connect these to the tree [7, 10] might improve performance. Note that such schemes require vastly more computation time. Another useful extension would be to use more elaborate rules for accepting voxels, fronts and segments, based on more

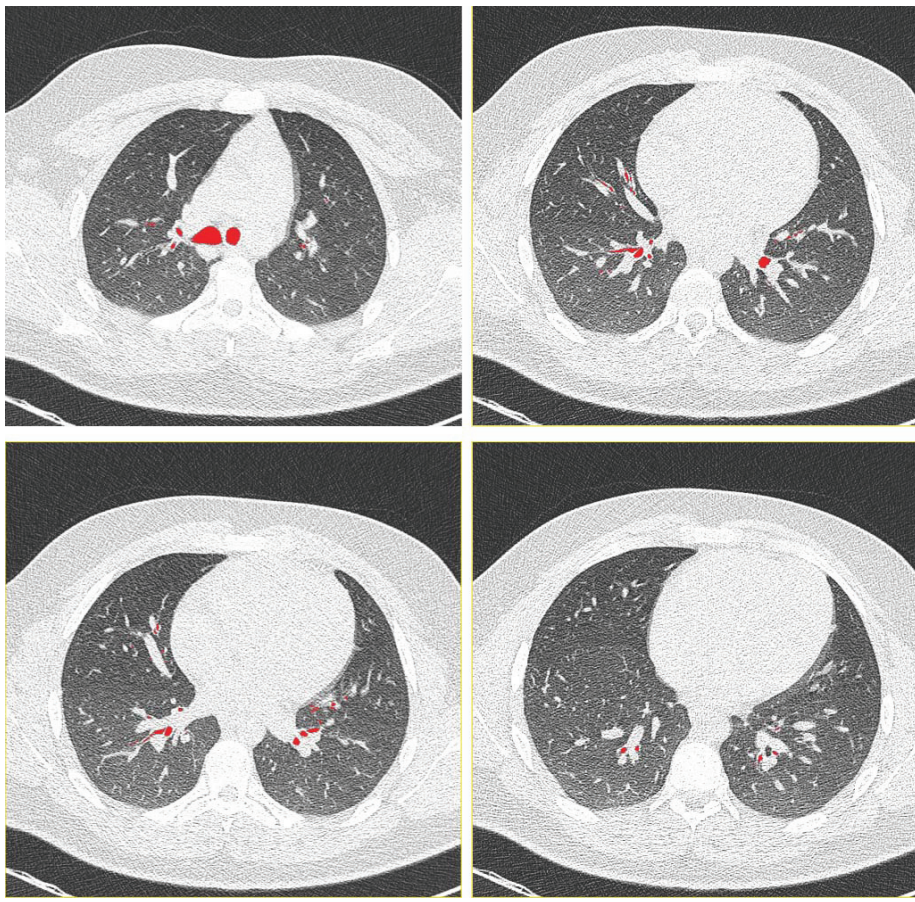


Fig. 5. Four axial slices for the case for which the highest percentage of branches were detected (case 21).

complex image information and statistical analysis, as was recently proposed in [11].

In conclusion, we have presented a fast, fully automatic method to extract the bronchial tree. The results on the EXACT09 data show that the method is able to find a large number of correct airway branches at the cost of a high average leakage volume.

References

1. Sluimer, I.C., Schilham, A.M.R., Prokop, M., van Ginneken, B.: Computer analysis of computed tomography scans of the lung: a survey. IEEE Transactions on Medical Imaging **25**(4) (2006) 385–405

2. Mori, K., Hasegawa, J., Suenaga, Y., Toriwaki, J.: Automated anatomical labeling of the bronchial branch and its application to the virtual bronchoscopy system. IEEE Transactions on Medical Imaging **19**(2) (2000) 103–114
3. Swift, R.D., Kiraly, A.P., Sherbondy, A.J., Austin, A.L., Hoffman, E.A., McLennan, G., Higgins, W.E.: Automatic axis generation for virtual bronchoscopic assessment of major airway obstructions. Computerized Medical Imaging and Graphics **26** (2002) 103–118
4. Schlathölter, T., Lorenz, C., Carlsen, I.C., Renisch, S., Deschamps, T.: Simultaneous segmentation and tree reconstruction of the airways for virtual bronchoscopy. In: Proceedings of the SPIE. Volume 4684. (2002) 103–113
5. Kiraly, A.P., Helferty, J.P., Hoffman, E.A., McLennan, G., Higgins, W.E.: Three-dimensional path planning for virtual bronchoscopy. IEEE Transactions on Medical Imaging **23**(11) (2004) 1365–1379
6. Bülow, T., Lorenz, C., Renisch, S.: A general framework for tree segmentation and reconstruction from medical volume data. In: Medical Image Computing and Computer-Assisted Intervention. Volume 3216 of Lecture Notes in Computer Science. (2004) 533–540
7. Fetita, C.I., Prêteux, F., Beigelman-Aubry, C., Grenier, P.: Pulmonary airways: 3-D reconstruction from multislice CT and clinical investigation. IEEE Transactions on Medical Imaging **23**(11) (2004) 1353–1364
8. Tschirren, J., Hoffman, E.A., McLennan, G., Sonka, M.: Intrathoracic airway trees: segmentation and airway morphology analysis from low-dose CT scans. IEEE Transactions on Medical Imaging **24**(12) (2005) 1529–1539
9. Palágyi, K., Tschirren, J., Hoffman, E.A., Sonka, M.: Quantitative analysis of pulmonary airway tree structures. Comput Biol Med **36**(9) (2006) 974–996
10. Higgins, W.E., Graham, M.W., Gibbs, J.D.: Robust system for human airway tree segmentation. In: Proceedings of the SPIE. Volume 6914. (2008)
11. Lo, P., de Bruijne, M.: Voxel classification-based airway tree segmentation. In: Proceedings of the SPIE. Volume 6914. (2008)
12. Tschirren, J., McLennan, G., Palágyi, K., Hoffman, E.A., Sonka, M.: Matching and anatomical labeling of human airway tree. IEEE Transactions on Medical Imaging **24** (2005) 1540–1547
13. Bülow, T., Lorenz, C., Wiemker, R., Honko, J.: Point based methods for automatic bronchial tree matching and labeling. In: Proceedings of the SPIE. Volume 6143. (2006)
14. Wiemker, R., Ekin, A., Opfer, R., Bülow, T., Rogalla, P.: Unsupervised extraction and quantification of the bronchial tree on ultra-low-dose vs. standard-dose CT. In: Proceedings of the SPIE. Volume 6143. (2006)
15. van Ginneken, B., Baggerman, W., van Rikxoort, E.M.: Robust segmentation and anatomical labeling of the airway tree from thoracic CT scans. In: Medical Image Computing and Computer-Assisted Intervention. Volume 5241 of Lecture Notes in Computer Science. (2008) 219–226
16. van Rikxoort, E.M.: Segmentation of Anatomical Structures in Chest CT Scans. PhD thesis, Image Sciences Institute, Utrecht, The Netherlands (2009)
17. van Rikxoort, E.M., de Hoop, B., Viergever, M.A., Prokop, M., van Ginneken, B.: Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection. Medical Physics **36**(7) (2009) 2934–2947

VOLCANO '09 Challenge Papers

The VOLCANO'09 Challenge: Preliminary Results

Anthony P. Reeves[1], Artit C. Jirapatnakul[1], Alberto M. Biancardi[1], Tatiyana V. Apanasovich[2], Chris Schaefer[3], Jeffrey J. Bowden[3], Markus Kietzmann[4], Rene Korn[4], Markus Dillmann[4], Qiang Li[5], Jiahui Wang[5], Jan H. Moltz[6], Jan-Martin Kuhnigk[6], Tatsuro Hayashi[7], Xiangrong Zhou[7], Hiroshi Fujita[7], Thomas Duindam[8], Bram van Ginneken[8], Rick Avila[9], Jane P. Ko[10], Kira Melamud[10], Henry Rusinek[10], Rafael Wiemker[11], Grzegorz Soza[12], Christian Tietjen[12], Matthias Thorn[12], Michael F. McNitt-Gray[13], Yanisley Valenciaga[13], Maryam Khatonabadi[13], Yoshiki Kawata[14], Noboru Niki[14]

[1]School of Electrical and Computer Engineering, Cornell University, Ithaca (NY), USA
[2]Jefferson Medical College, Thomas Jefferson University, Philadelphia (PA), USA
[3] Imaging Services, Biomedical Systems, USA
[4] Definiens AG, Germany
[5] Department of Radiology, Duke University Medical Center, USA
[6] Institute for Medical Image Computing, Fraunhofer MEVIS, Germany
[7] Gifu University, Japan
[8] Image Sciences Institute, University Medical Center Utrecht, the Netherlands
[9] Kitware, Inc.
[10] Radiology Department, New York University, USA
[11] Philips Research Hamburg
[12] Computed Tomography, Healthcare Sector, Siemens AG, Germany
[13] David Geffen School of Medicine at UCLA, USA
[14] University of Tokushima

reeves@ece.cornell.edu, tatiyana.apanasovich@jefferson.edu, cschaefer@biomedsys.com, mkietzmann@definiens.com, jiahui.wang@duke.edu, jan.moltz@mevis.fraunhofer.de, hiroshi.fujita.gifu@gmail.com, bram@isi.uu.nl, rick.avila@kitware.com, hr18@nyu.edu, rafael.wiemker@philips.com, grzegorz.soza@siemens.com, mmcnittgray@mednet.ucla.edu, niki@opt.tokushima-u.ac.jp

Abstract. The VOLCANO'09 Challenge invited participants to evaluate the change in size of pulmonary nodules in CT images; the challenge data set consisted of 50 pairs of CT scans each scan containing a single nodule. This is the first challenge for CAD methods on pulmonary nodules in which size change rather than volume estimation is the primary endpoint. Responses from 13 teams were received with size change results for a total of 17 different methods. In this paper the challenge data set is described and statistical results computed from the submissions are presented. The dataset consisted of several subgroups: (a) zero-change cases, cases with different slice thickness scans, cases with actual size change and a synthetic nodule case. No statistical difference was found between the methods; a slice thickness change was significant and there was an interesting bias observed for some zero-change nodules.

Keywords: Size Change Estimation, Variability Analysis, Pulmonary Nodules.

1 Introduction

The target of the challenge is three-dimensional change analysis of pulmonary nodules in CT images. The focus of the challenge is not directly on image segmentation itself (which tells us little of the underlying disease) but rather the change in size of the nodule recorded on two time-separated images. This size change is a critical measurement for (a) diagnosing cancer and (b) evaluating response to therapy. One of the most important indicators of malignancy is the relative change in size of a nodule over a period of time. The critical issue for the challenge, the precision of size change measurement, is needed to establish the minimum time delay between sequential scans and the associated magnitude of the measurement required to determine malignancy or response to therapy. There has been one previous pilot study in this area, Biochange 2008 [1].

Most evaluation methods for CAD systems, including challenges, involve a ground truth established be experts. However, for the task of nodule size estimation it is well known that there is a large amount of variation or disagreement in expert size estimations [2]. Further, it has not been established that expert's manual estimations are superior to automated measurements. In this challenge, while the change in size of nodules will be reviewed by experts, we explore the issue of ground truth through the submitted responses to the challenge.

1.1 Motivation for the study

Current approaches to quantification of nodule volume change measurement exhibit two main problems that complicate their direct comparison. First, these methods require a large unified database of both stable and growing nodules. Second, there is no single commonly used evaluation technique that would assess the measurement quality of a particular method. Therefore we invited interested parties to take part in this unique study that addresses both of these issues by providing a single evaluation image dataset and a common methodology for assessing the quality of the measurement algorithm.

1.2 The VOLCANO Challenge

The challenge involved measuring the change in nodule size for 50 scan pairs. Four additional scan pairs were made available for training. Teams reported the fractional change in nodule size for each of the 50 scan pairs. Thirteen different teams submitted their measurement change results from on a total of 17 different methods. In 12 of these methods, the actual volumes recorded for each nodule were also reported.

The participants were only informed that there were 50 nodule pairs; however, the data may be divided into four subgroups:

A. (14) zero-change in which the scans were taken minutes apart and therefore there is no real change in the nodule size.
B. (13) zero-change cases as in A above except one scan had a slice thickness of 1.25 mm and the second scan had a larger slice thickness (2.5 or 5.0 mm)

C. (19) nodules with a significant time interval between scans and therefore some real change and (3) nodules with a large change in size of greater than 150%, one of which was known to be malignant. Of these nodules, 19 were considered to be stable or benign by biopsy and 3 were diagnosed as malignant.
D. (1) synthetic phantom nodule with a known size recorded multiple times with different slice thicknesses

In general, the main interest is to learn the smallest size change which can be reliably detected and the precision in that size change measurement. A number of studies on repeat scans have been reported in the literature ([3] refers to several of these). In these studies the limits of agreement for repeat scans of the same nodule are in the order of 20-25% by volume. For theses reasons most of the cases in data set C were selected to have a size change within the range of ±50%. For completeness, three cases with a very large size change (150% or more) were included to characterize the measurement methods for such situations.

1.3 Overview

In this paper we report the initial statistical findings for the data submitted by the VOLCANO teams that provides collective information about the nodule size change measurement process; we do not provide a performance rating for the different teams. This paper addresses the following issues:

1. What is the precision of change measurement to be expected from computer assisted measurement methods? This question is addressed by considering the size change measurements from group A.
2. What is the impact on the computer methods of changing the slice thickness of the CT scan? This is addressed by the analysis of any bias for the size change measurements of group B and a comparison of the variation of groups A and B. Also, we can observe if there is a bias in group D.
3. What is the variation to be expected between different computer methods when there is an actual size change? This is addressed by the variation in measurements for group C.
4. For the teams that provided volume information, what is variation in volume estimates made by the different methods? For comparative volume measurements we consider groups A-C and for absolute volume measurement we consider group D.

2 Materials and Methods

The image data used in the study was acquired for the Public Lung Database to address drug response [4] and was provided by the Weill Cornell Medical College with the exception of one case of a synthetic “phantom” nodule provided by the FDA [5]. Cases were selected that contained at least one nodule of solid consistency which was present in at least two scans with a whole-lung field of view; only nodules visible on at least three slices on both scans were included.

Table 1. Summary of scan parameters

Group	Current (mA)	kVp (kVp)	Scanner Models
A	40-250	120	GE LightSpeed (Ultra, QX/i, Pro 16)
B	20-80	120	GE LightSpeed (Ultra, QX/i)
C	40-80	120	GE LightSpeed (Ultra, Pro 16, VCT)
D	40	120	Philips MX8000 IDT 16

The VOLCANO'09 dataset consists of 50 pairs of CT scans of pulmonary nodules: 49 scan pairs of real pulmonary nodules and one phantom nodule. The data consisted of four subgroups A-D as outlined in the introduction. The size distribution for the nodules used in this study is shown in Figure 1.

The scan pairs in group A had the same slice thickness for both scans; in 13 cases, both scans had a slice thickness of 1.25 mm while one case had scans with a slice thickness of 2.5 mm. For group B, 11 cases had scans with 1.25 mm and 2.5 mm slice thickness, while two cases had scans of 1.25 and 5.0 mm slice thickness. All the scans in group C had a slice thickness of 1.25 mm. Scans in these three groups were acquired without overlap. Other relevant scan parameters are listed in Table 1.

For group C, the status of the nodule was determined by a radiologist; stable nodules were either biopsied (3) or did not have any clinical change in 2 years (16), while the three malignant nodules were biopsied.

Group D was comprised of the synthetic phantom nodule, a 10 mm (523.60 mm^3) sphere with two different slice thickness reconstructions, 1.5 mm and 3.0 mm with 50% overlap. All scans have a whole-lung field of view. The phantom was placed in a chest phantom with simulated vascular structures [5].

In five zero-change cases, the patient was oriented in a different position between the scans; one case was in group A while four cases were in group B.

2.1 VOLCANO Data Preparation

Prior to making the images available for the challenge, all identifying patient information was removed. The original dates of the scans were replaced with dates corresponding to a time interval of 100 days between scans, with the order of the scans was randomized. The scans were then clipped in the axial direction, with the five slices below and above the region containing the nodule included if possible. This was done (a) because some of the scans did not cover the whole-lung in the axial dimension and (b) to reduce the amount of data to be downloaded for this study.

Along with the scans, teams were provided with a spreadsheet identifying the approximate center of the nodule established by a human observer.

2.2 The data measurement methods

Submissions to the VOLCANO challenge were received from 13 teams. Several teams submitted multiple size change measurement methods for a total of 17 submitted methods. Although not required, 12 teams provided volume measurements for each nodule. The teams and their methods are summarized in Table 2.

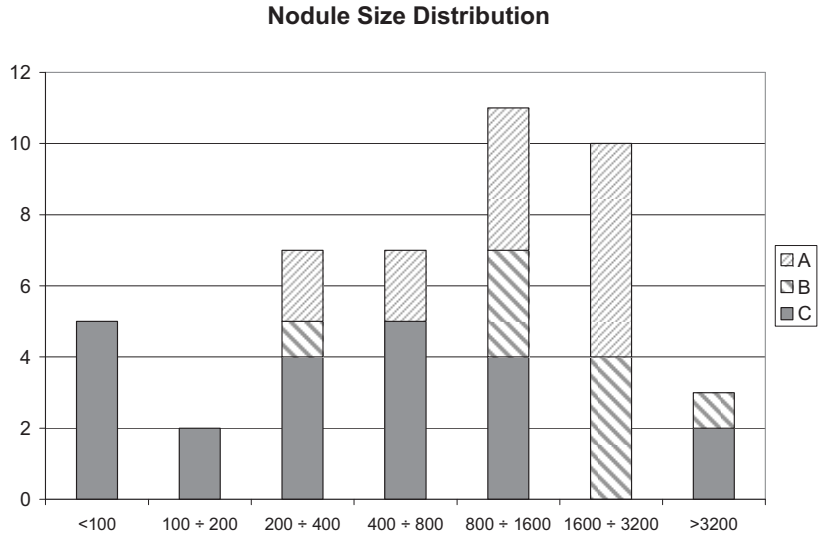


Fig. 1. The distributions of the approximate nodule sizes, expressed as volumes in mm^3, according to the different groups they belong to. The listed volumes (for which an equivalent diameter is defined as the diameter of the sphere having that volume) correspond to the following equivalent diameters, respectively: 5.8, 7.3, 9.1, 11.5, 14.5, and 18.3 mm.

Table 2. The methods submitted to VOLCANO'09. (PVC = partial voxel compensation)

Method	Team/Method	Automation	Method
1	Tokushima	3	Image filtering
2	ISI, SPHERE	4	Sphere fitting
3	ISI, SEG	4	Image filtering
4	ISI, REG	4	Elastic registration
5	NYU, HYB	7	Image filtering and PVC
6	NYU, HYBA	7	Image filtering and PVC
7	UCLA	6	Image filtering
8	VIA, GAD	1	Density change
9	VIA, GAS	4	Image filtering
10	Kitware	4	Fast marching and shape detection level set
11	Duke	1	Spiral scanning, dynamic programming
12	Gifu	1	Image filtering
13	Biomedsys	2	Image filtering
14	MeVIS	3	Image filtering and PVC
15	Siemens	3	Image filtering and PVC
16	Philips	1	Active contour
17	Definiens	5	Image filtering

The methods had varying levels of operator interaction which can be divided broadly into three groups: completely automated after specification of a seed point, manual parameter control, and modification of the resulting boundary or indicating additional control points. These categories were further subdivided according to the fraction of cases which required manual intervention. Teams were asked to rank the level of automation required by their algorithms using the following scale:

1. Totally automatic using seed points
2. Limited parameter adjustment (on less than 15% of the cases)
3. Moderate parameter adjustment (on less than 50% of the cases)
4. Extensive parameter adjustment (more than 50% of the cases)
5. Limited image/boundary modification (on less than 15% of the cases)
6. Moderate image/boundary modification (on less than 50% of the cases)
7. Extensive image/boundary modification (more than 50% of the cases)

The level of automation required for each method is detailed in Table 2.

Although each method was unique, there were several common approaches to the task of volume change measurement. One approach taken by seven methods (Tokushima, ISI-Seg, UCLA, VIA-GAS (Vision and Image Analysis Group, Cornell), Gifu, Biomedsys, Definiens) is based on simple image filtering operations [6]. Generally, methods using this approach extracted a volume of interest (VOI) around the seed point for each nodule. Definiens additionally used an automatically generated anatomical model to further restrict the VOI. This volume of interest was resampled and a threshold applied to identify voxels belonging to high-intensity structures. Next, either region growing or connected component analysis is applied to the volume of interest to eliminate non-nodule structures, followed by the removal of attached structures such as vessels or the chest wall using morphological filtering or other more advanced techniques. Four methods (MeVIS, NYU-HYB, NYU-HYBA, Siemens) extended this approach to better address partial voxels along the perimeter of the nodule [7, 8]. These methods determine a region around the border of the nodule where the voxels have intensity between that of solid tissue and the lung parenchyma; based on histogram analysis, these voxels are weighted when computing the nodule volume.

The remaining six methods used different approaches. Most methods resampled the CT scans into isotropic space. ISI-Sphere estimated the best fitting spherical volume of interest at the seed point of the nodule from a thresholded, resampled volume of interest. The volume of the nodule was estimated from the number of voxels above a threshold. The ISI-Registration method applied non-rigid registration to deform the first scan to the second; this transformation was then applied to a segmentation obtained for the first scan to obtain the volume of the nodule on the second scan. Kitware required only a manually specified seed point and a bounding box. Their method computed several features for each voxel, including lung wall, vesselness, gradient, and intensity features which were aggregated and used to guide a fast marching algorithm to generate an initial guess of the nodule boundary; this guess was refined using a shape detection level set. The volume was computed from the surface of the resulting level set. Duke used a spiral-scanning technique to convert the 3D volume of interest around the nodule to a 2D generalized polar coordinate system. Dynamic programming techniques were used to obtain the nodule boundary on the 2D image which was then transformed back into 3D space [9]. This boundary was

applied to the original 3D image to estimate the nodule volume. Philips used an active contour-based approach utilizing a radial basis function energy minimization algorithm [10]. In contrast to the methods described thus far, VIA-GAD (Cornell) did not explicitly segment the nodule; instead, the change in nodule size was estimated from the change in density of a Gaussian-weighted region around the nodule [11].

2.3 Data analysis

Teams were requested to provide a size change metric for each nodule. For methods using volumes, the size change metric would be:

$$RVC = \frac{V_2 - V_1}{V_1} \quad (1)$$

The data metric that was requested for results (relative change with respect to time 1) is asymmetric with respect to the order of presentation: no change in size has a value of 0, a 10 times increase in size results in a value of 9 while a 10 times decrease in size results in a value of - 0.9. The negative size change is bounded by -1 while the positive size change is unbounded. The order of the scan pairs was randomized. It is possible to reverse the order: that is computing the relative size change from scan two to scan one by the following transformation:

$$RVC' = \frac{1}{RVC + 1} - 1 = \frac{V_1}{V_2} - 1 \quad (2)$$

where RVC represents the reported relative size change. However, if any of the methods have an order bias, flipping the results to match the correct time sequence would mask such an effect. The one data set where ordering is important is data set B; we flipped the results so that the thin slice scan was always the first in the pair to determine the effect caused by a change in slice interval. For the other data sets flipping to restore the correct temporal order of the scans was not performed.

To quantify the variation and bias of each group, the median of the median of absolute deviation (MMAD) and median of the absolute median (MAM) were computed for each group. The median of the absolute deviation (MAD) is computed by taking the median size change metric for each nodule, and, for each method, computing the absolute deviation from the median. This results in a MAD value for each nodule; the median of these is then reported as the MMAD. The MAM is computed by taking the median of the absolute median for each nodule. Both these measures can be computed for each method instead of for each nodule.

The statistical agreement between methods was established for the size change measurements using the Friedman test. Groups were compared by applying the Wilcox rank-sum test to the MAD values computed for each nodule.

3 Results

Box and whisker plots for the different groups are shown in Figures 2 to 6. In the plots, the median of each nodule is indicated by a thick line. The 25% to 75% interquartile range (IQR) is represented by the box. Methods with values inside the box are generally in good agreement; 50% of the methods lie inside the IQR. The lines above and below the box ("whiskers") represent the largest and smallest values that are within 1.5 times the IQR; any more extreme values are considered outliers and indicated by an open circle. The number shown x-axis is a random case identifier that does not correspond to the case identifiers given to the teams; the results are ordered according to median change. For group C, the three cases with the most change are plotted separately for visibility.

For each of the groups, the median of the absolute medians (MAM) and the median of the median of the absolute deviations (MMAD) were computed across methods. These are measures of the bias and variation, respectively, and reported in Table 3.

Table 3. Summary of the median absolute median (MAM) and median of median absolute deviation (MMAD) of the relative size change for each group. Note that the results of group D are on a single nodule.

Group:	A	B	C	D
MAM	0.0572	0.1020	0.1740	0.0620
MMAD	0.0388	0.0899	0.0592	0.0761

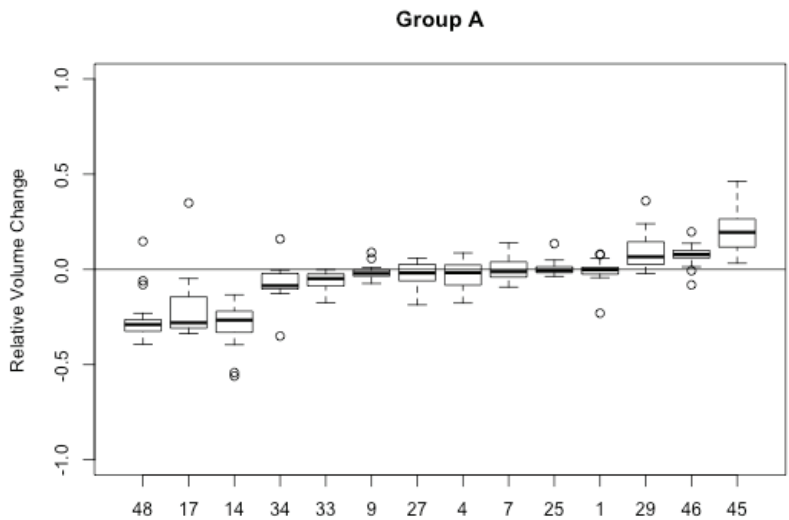


Fig. 2. Box plot for group A, the zero-change dataset

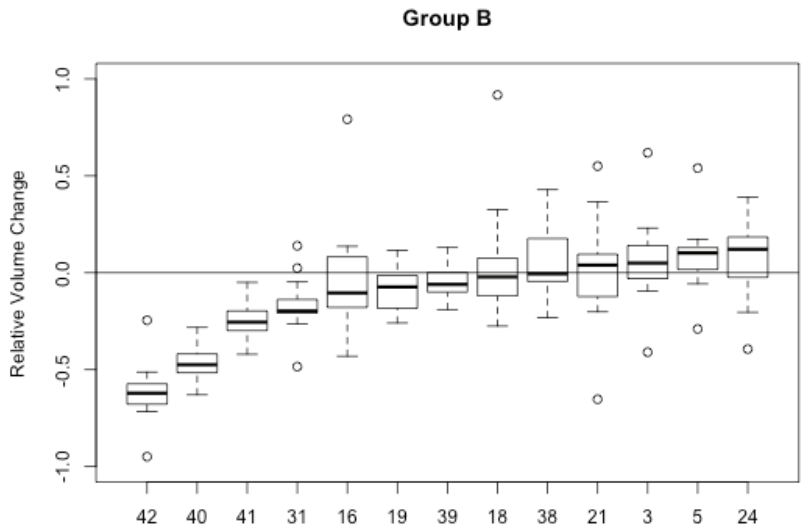


Fig. 3. Box plot for group B, zero-change with different slice interval, with the volume change inverted when necessary to order by small to large slice interval. All nodules were imaged on scans of 1.25 and 2.5 mm slice thickness except for 3 and 21 which were 1.25 and 5.0 mm.

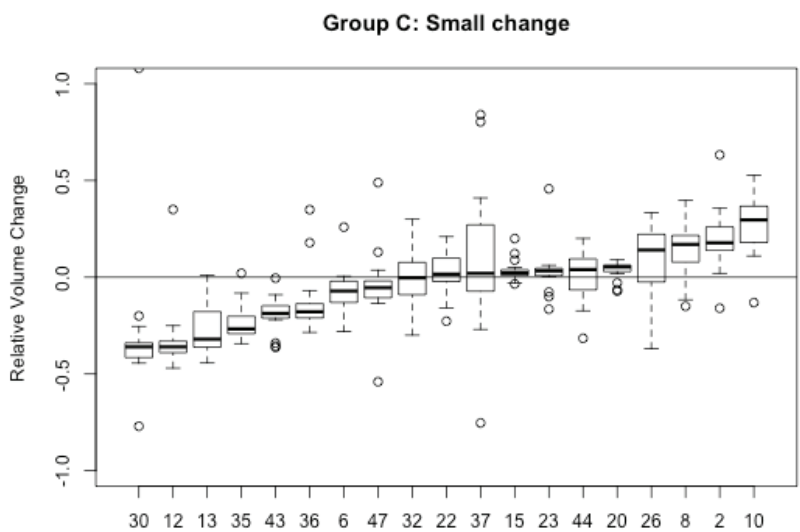
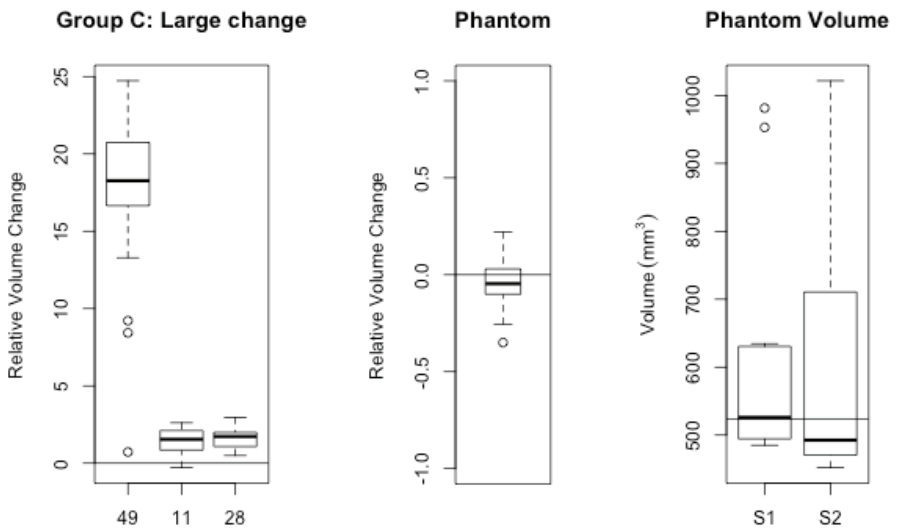


Fig. 4. Box plot for group C, the actual small change nodules. One outlier is not shown.



(a) (b) (c)

Fig. 5. (a) Group C nodules with large change, with the volume change for the first nodule inverted for visibility and (b) size change and (c) volume variation results for the synthetic nodule for scan 1 (S1) and scan 2 (S2). The volume results for the phantom are presented for 12 of 17 methods that provided volumes.

4 Discussion

4.1 The repeat measurement behavior of the computer assisted methods

In group A, the scans are a repeat measure on nodules that have not changed in size. The median MAD was 0.0387, which indicates the level of variation between the methods. The bias from the true zero value of size change is represented by the median absolute median of 0.0572.

To examine the variation due to the nodule and/or scanner, the median MAD was computed across nodules. The value of 0.0700 was larger than the median MAD computed across methods; thus, the variation due to the nodule/scanner is larger than the variation due to the methods. In addition, the median of the absolute medians was lower, 0.0217, suggesting that any bias that exists is due to the nodule/scanner and not the method. The 85% confidence interval of the absolute medians across methods (i.e., omitting the two largest cases) is 0.266 or a 26% volume change. It is interesting to note that for 8 of the 14 cases the interquartile range (IQR) does not include zero. Therefore there is evidence of a systematic bias introduced by the scanner/nodule combination for these cases.

4.2 The impact of change in the CT scan slice thickness

The size change measurements in group A for all the methods are in agreement ($p=0.92$ using the Friedman test). In contrast, the size change measurements for group B, where there is a change in slice thickness, are significantly different between the methods ($p < 0.01$) according to the Friedman test. To determine if the variation is greater in group B, the median absolute deviation (MAD) was computed for each nodule in group A and group B and the Wilcoxon rank-sum test was performed. The variation was higher in group B than group A ($p < 0.01$); with $p=0.05$, the variation is 40% higher in group B than group A.

We observe from Fig. 5c that there is a greater dispersion in the IQR of the volume measurement for the thicker slice scan of the phantom nodule (131.01 mm^3 vs. 207.79 mm^3) which is consistent with the above.

4.3. The variation of the methods when there is a real size change

The variation with respect to no-change and real change data sets is achieved by comparing the median of the median of absolute deviation (MAD) values between group A and the 19 small change cases in group C. The median MAD for group A and the small change cases in group C were 0.0387 and 0.0590 respectively. The small increase might suggest that the results obtained from a zero-change dataset capture similar behavior of a dataset with a small amount of change as represented by group C. Furthermore, comparison with the 3 nodules with larger size changes in group C showed that there was an increased variation (median MAD of 0.4446) for a larger size change.

4.4 A comparison of volume estimation and size change measurement

For the 11 teams that provided volume information, the variation in the size change measurement was compared to the volume measurements for group A and the small change subset of group C. The Friedman test was performed to determine if the methods were in agreement.

In group A, there was no significant difference between the size change measurements of the methods ($p = 0.92$), but there was significant disagreement between the volume measurements of the methods ($p < 0.01$). For the small change subset of group C, again there was no significant difference between the size change measurements of the methods ($p = 0.97$), but significant disagreement in the volume measurements ($p < 0.01$). This may be due to a bias in volume measurement between methods that might be neutralized when computing size change.

5 Conclusion

Change in size measurements made on 50 nodule image pairs were reported for 17 different methods. The analysis of the results showed (a) that there was no statistical difference between the methods on scans of the same slice thickness, (b) that there was a statistical difference in the methods when the scan slice thickness is changed,

and (c) that the behavior of the methods for nodules with a small real change in size was similar to that for the zero-change data. The last point has implications for the validity of using zero-size change datasets for evaluating nodule measurement performance. For 11 of the methods volume measurements were provided in addition to the size change measurements. The volume measurements did show a statistical difference between methods; therefore, caution is needed when extrapolating from studies that focus only on volume estimation when size change is the intended task.

Acknowledgments

We are grateful to David Yankelevitz of the Weill Medical College of Cornell University for reviewing the cases used in this study. We gratefully acknowledge Lisa Kinnard of the FDA for providing us with the synthetic nodule CT images used in this study.

References

1. Fenimore, C. Biochange 2008 Pilot, http://www.itl.nist.gov/iad/894.05/biochange2008/Biochange2008-webpage.htm
2. Reeves, A. P., Biancardi, A. M., Apanasovich, T. V. et al.: The Lung Image Database Consortium (LIDC): A Comparison of Different Size Metrics for Pulmonary Nodule Measurements. Academic Radiology 14, 1475--1485 (2007).
3. Gavrielides M.A., Kinnard L.M., Myers K.J., Petrick N.: Noncalcified lung nodules: volumetric assessment with thoracic CT. Radiology 251, 26--37 (2009).
4. Reeves, A. P., Biancardi, A. M., Yankelevitz, D. F., Fotin, S., Keller, B. M., Jirapatnakul, A. C., Lee, J.: A Public Image Database to Support Research in Computer Aided Diagnosis, In: 31st Annual International IEEE EMBS Conference, in press, IEEE Press, New York (2009).
5. Gavrielides, M.A., Zeng, R., Kinnard, L.M., Myers, K.J., Petrick, N.: A model-based approach for the analysis of lung nodules in a volumetric CT phantom study. In: SPIE Medical Imaging Conference, vol. 7260, pp. 726009, SPIE (2009).
6. Kostis, W. J., Reeves, A. P., Yankelevitz, D. F., Henschke, C. I.: Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images. IEEE Trans. on Medical Imaging 22, 1259--1274 (2003).
7. Ko, J.P., Rusinek, H., Jacobs, E.L., Babb, J.S., Betke, M., McGuinness, G., Naidich, D. P.: Small Pulmonary Nodules: Volume Measurement at Chest CT: Phantom Study. Radiology 228, 864--870 (2003).
8. Kuhnigk, J., Dicken, V., Bornemann, L., Bakai, A., Wormanns, D., Krass, S., Peitgen, H.: Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic CT scans. IEEE Trans. on Medical Imaging 25, 417--434 (2006).
9. Wang, J., Engelmann, R., Li, Q.: Segmentation of pulmonary nodules in three-dimensional CT images by use of a spiral-scanning technique. Medical Physics 34, 4678--4689 (2007).
10. Opfer R., Wiemker R.: A new general tumor segmentation framework based on radial basis function energy minimization with a validation study on LIDC lung nodules. In: SPIE Medical Imaging Conference, vol. 6512, pp. 651217, SPIE (2007).
11. Jirapatnakul, A. C., Reeves, A. P., Biancardi, A. M., Yankelevitz, D. F., Henschke, C. I.: Semi-automated measurement of pulmonary nodule growth without explicit segmentation. In: IEEE International Symposium on Biomedical Imaging. pp. 855-858. IEEE Press, New York (2009).

Algorithms, Architecture, Validation of an Open Source Toolkit for segmenting CT Lung lesions

Karthik Krishnan[1], Luis Ibanez[1], Wesley D. Turner[1], and Ricardo S. Avila[1]

Kitware Inc., 28, Corporate Drive, Clifton Park, NY, USA.
karthik.krishnan@kitware.com,

Abstract. We implented a lesion segmentation toolkit with a focus on segmenting solid and part-solid lesions from lung CT. The algorithm constructs four three-dimensional features to detect the lung wall, vessels, soft and diffuse tissue and edges. These features form boundaries and propagation zones that guide the evolution of a subsequent shape detection level set. User input is used to determine an initial seed point for the level set and users may also set a region of interest around the lesion. The methods are validated against 18 nodules on an anthropomorphic thorax phantom simulating a realistic anatomy, acquired under various scanner parameters. We also validated repeatability using 6 coffee-break (zero volume change) clinical cases. The methods, clinical and testing datasets have been made available to encourage reproducibility and foster scientific exchange. We integrated the methods into a user-friendly visualization application. We will also describe how one may obtain the toolkit and generate results for the VOLCANO workshop.

1 Introduction

Computed Tomography (CT) is the preferred modality for the detection of cancerous lung lesions. Quantifying tumor size and changes in size over time is necessary to evaluate the effectiveness of a drug. Tumor size is currently quantified using the uni-dimensional RECIST [1] [2] or bi-dimensional WHO [2] [3], both of which use lesion diameter as surrogates for volume to determine an effective tumor measurement. Neither of these methods take the actual volume of the tumor into account. With the advent of MSCT, volumetric segmentation of lung nodules allows for more accurate quantification of the tumor burden and better determination of changes in tumor burden over time and in response to drug regimen.

The goal of this paper is to describe the algorithm and the architecture of an open source lesion sizing toolkit. We will validate it on phantoms as well as clinical data. We will also describe how one may obtain the toolkit and generate results for the VOLCANO [4] workshop. The rest of the paper is organized as follows. Section 2 summarizes the segmentation methods employed. Section 3 summarizes the validation of the methods on synthetic phantoms having known volumetric ground truth from the FDA. Section 4 outlines the architecture of the toolkit and describes the use of the toolkit from VolView, a commercially

available data visualization application. Integration with VolView allows users of the toolkit to load and visualize datasets as well as to perform volumetric analysis in an intuitive and powerful graphical user interface. Finally, in Section 5 we summarize our performance on the Workshop evaluation dataset and provide the user with those resources necessary to duplicate our results.

2 Lesion Segmentation

Lung nodules exhibit a wide range of morphologies and shapes. They may be solid, part-solid (solid nodules with a diffuse region) or non-solid. Solid nodules have a density that is close to soft tissue while non-solid nodules have a density between air and soft tissue. Nodules are generally attached to vessels and often to the lung wall or the mediastinum. Coding a single algorithm that effectively segments a nodule under these conditions is difficult. We choose to take a modular and extensible approach that combines a region growing algorithm - in this case a shape detection level set - with a set of feature generation filters. The region growing algorithm is allowed to evolve from a user provided seed under the aggregate guidance of the features. Our current implementation uses 4 features; to account for the lung wall, vessels, gradients and tissue intensities (soft and diffuse). Together, these features allow propagation of the segmentation front within the intensity thresholds specified by the intensity feature while detering propagation into the lung wall, into vessels, or across sharp edges. The addition of new features to specifically account for areas such as eg. the hilar region is easily supported.

2.1 Feature generation

Prior to any feature generation, the CT dataset is preprocessed by resampling to isotropic resolution matching the smallest voxel spacing. This is followed by generation of various features:

Lung wall feature The lung wall is segmented by thresholding followed by curvature constrained front propagation. The dataset is binarized so as to retain voxels above -400HU. We then run hole filling on the resulting binary image. The hole filling operation uses a quorum (voting) algorithm to decide whether a new pixel will be *filled* at every iteration. Changing the voting threshold is essentially equivalent to changing the maximum curvature allowed in the final binary segmentation. We use a $7\times7\times7$ manhattan kernel with a voting threshold of 1 (lowest possible curvature). The filter is run interatively until no voxel changes status. Fig.1 shows the evolution of the front. Admittedly, the final contour is slightly concave rather than convex, since the quorum voting does not distinguish between positive and negative curvature, but the wall detection recovers most of the nodule without allowing bleeding into the lung wall.

The resulting image is transformed using a sigmoid with inversion, resulting in a grayscale representation of the inner lung, with a gradual rolloff towards the lung wall.

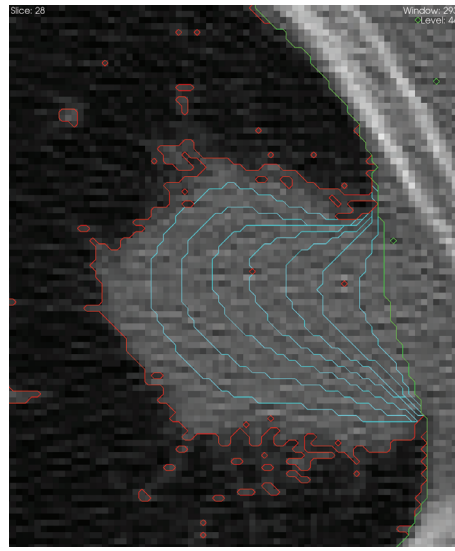


Fig. 1. Segmentation of the lung wall by voting based front propagation on a lesion attached to the lung wall. The front starts with the red contour (obtained by thresholding the dataset at -400HU. In cyan are contours of the front at iterations of 10, 20, 30, 40, 50, 60, 80. The algorithm converges in 151 iterations at the green front.

Vesselness feature Vesselness features are computed using Sato [5] tubularness. ($\alpha = 0.1$, $\beta = 2.0$). The resulting image is transformed using a sigmoid filter with inversion resulting in a grayscale image with a rolloff towards tubular structures.

Gradient feature The gradient feature attempts to localize the edges of the lesion. This is modeled using a Canny edge detector [6], which achives optimal edge localization (with voxel accuracy). The result is a voxelized edge map. The resulting image is transformed using a sigmoid with inversion resulting in a grayscale image with a rolloff towards the edges.

Intensity feature The intensity feature is based on a simple thresholding of the dataset, followed by transformation with a sigmoid filter to generate gradual rolloffs. The user specifies the type of the lesion (solid / part-solid), which determines the threshold level: -200HU for solid lesions and -500HU for part-solid lesions.

Aggregation of features Conceptually, each feature image represents the likelihood of a particular voxel being part of the lesion from a single perspective. Fig 2 and 4 show the features on a solid and a part-solid lesion. An aggregation

Algorithm 1 Voting based front propagation

Let Kernel Radius $r = 3$
Let BirthThreshold, $T = r^3 + 1$. For a $7 \times 7 \times 7$ city-block kernel, $T = 171$
Let A_i be indices of voxels on the front at iteration i
Threshold CT image at -400HU. Voxels above are set to 1, below to 0.
Iterate over the image and add all indices within r voxels of the boundary to A_0
repeat
  $PixelsChangingStatus \Leftarrow 0$
  for each voxel in A_i **do**
    Check for quorum, Q. (Quorum is achieved if number of ON voxels within a neighborhood ($7 \times 7 \times 7$) centered at the current voxel $> T$.
    if Q is *true* **then**
      Scehedule this pixel for inclusion in foreground.
      $++PixelsChangingStatus$
      Add background voxels in neighborhood to front for next iteration, A_{i+1}.
    else
      Schedule this pixel for inclusion in background.
      Add this location to the front for the next iteration, A_{i+1}
    end if
  end for
  Update the status of scheduled voxels.
  $++i$
until $PixelsChangingStatus = 0$

of the features is obtained by normalizing each individual feature to lie within the range 0 to 1 and then choosing the minimum feature value at each voxel. The resulting image is passed onto the segmentation module.

2.2 Segmentation Module

The segmentation module runs a segmentation algorithm guided by a feature probablity image. The toolkit is equipped with a variety of segmentation modules, wrapped in a way that allows them to be plugged in as components to the lesion segmenter. In our user interface, the user identifies the lesion with a single seed point. We use fast marching [7], followed by refinement with a shape detection level set [8]. The fast marching front starts at the seed and proceeds outwards, solving the eikonal equation, its propagation guided solely by the aggregate feature image and an optional, user-defined, region of interest (ROI). The output is a time crossing map, with each voxel representing the time taken for the front to reach that voxel. We use a stopping time of 5s implying that the front stops propagating when all voxels on the front have an arrival time greater than 5. The fast marching front is conservative enough to ensure that the segmentation results lie within the lesion. This resulting time crossing image is renormalized to the range $[-4, 4]$, so as to be symmetric about 0 and passed onto the shape detection module.

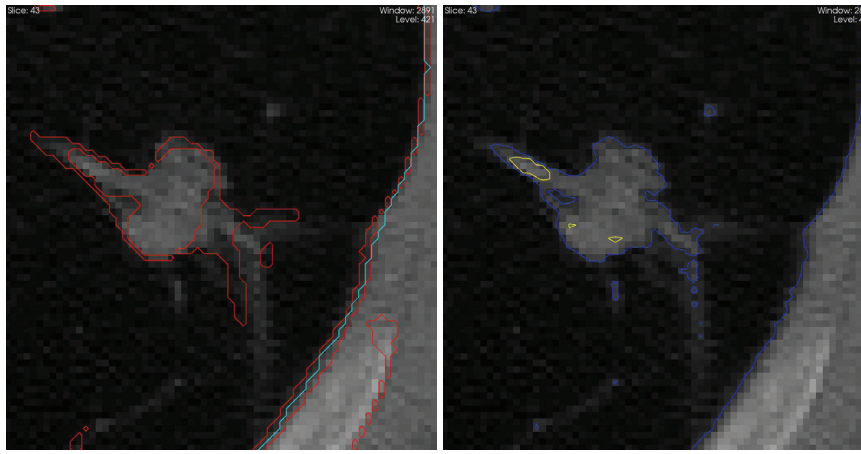


Fig. 2. Features (a) Lung wall features in cyan, Canny Edges features in red; a contour around the single voxel thick edges is shown (b) Intensity features in blue, vesselness features in yellow.

The shape detection level set is guided by the propagation and curvature components.

$$\phi_t + F(1 - wK)\,|\nabla\phi| = 0$$

The propagation F is the aggregate feature image. w is the curvature weight for the curvature, K. We use a propagation to curvature weight ratio of 500:1, to account for the wide variety of shapes that lesions can assume. The final lesion boundary is then taken as the -0.5 isosurface of the shape detection level set. Note that this isosurface value is driven by the characteristics of our aggregate feature map. In the nominal case of the lesion having a boundary within the parenchyma, we find that the canny edge is the dominant feature for lesion segmentation. Our canny edge feature only localizes the true edge boundary to a one voxel precision (Fig.4(b)). By taking the level set at -0.5 we force the segmentation to the interior of the canny detected boundary voxel and reduce our error bound from 1 to 0.5 voxels.

A surface of the resulting level set is generated using Marching Cubes [9] at an isovalue of -0.5. The volume is computed using a discretized version of the divergence theorem [10], to compute the volume in a closed triangle mesh.

3 Validation of the segmentation algorithm

3.1 Phantoms and Ground truths

We validated the results on an anthropomorphic thorax phantom [11][12] (courtesy US FDA), containing nodules of various densities, sizes, shapes and at various locations in the chest. This phantom provides ground truth for nodules

in a realistic anatomy (fig. 3) that includes attachments to vasculature, airways and the lung wall. Exposure varies from 20mAs to 200mAs. At the time of this writing, ground truth volumes are available only for the spherical nodules. We are awaiting ground truths via microCT on the spiculated, lobulated and elliptical nodules. The database consists of 6 CT scans, each containing a 5, 8 and 10mm diameter spherical nodule with a density of 100HU. Three of the scans have a slice thickness of 0.8mm and a z spacing of 0.4mm, the other three have a slice thickness of 3mm and a z spacing of 1.5mm.

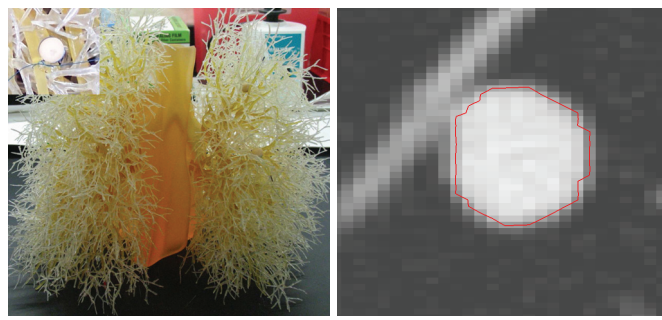


Fig. 3. *Left*: A photograph of the anthropomorphic phantom and an inlaid nodule (courtesy US FDA). *Right*: Its segmentation.

3.2 Statistics

Defining the error, $\xi = 100 \times \left|\frac{ComputedVolume - TrueVolume}{TrueVolume}\right|$, we find that for the 18 lesions in the FDA phantom, our average ξ was 30.24% (S.D 31.39%). The errors largely stemmed from a poor estimate of the features due to the resolution across slices. This was confirmed by noting that on the 9 lesions, with 0.8mm slice thickness, our average volumetric error, ξ was 11.75% (S.D 11.38%). Furthermore, the half voxel error (as mentioned above) causes an amplification of the errors for the small 5mm lesions. This is confirmed by noting that on the 6 lesions with a diameter $\geq 8mm$ and a slice thickness of 0.8mm, our average volumetric error ξ is 4.28% (S.D. 2.4%).

It is instructive to relate the error in volumetric estimate to the error in the uni-dimensional RECIST measurement. A 30% overestimate in a lesion with a diameter of 8mm is equivalent to a change in diameter of 0.73mm, Hence a ξ of 30% for a 8mm lesion would roughly which roughly translates to an error of a voxel. This is smaller than the intra as well as interobserver variability to perform a unidimensional tumor measurement [2].

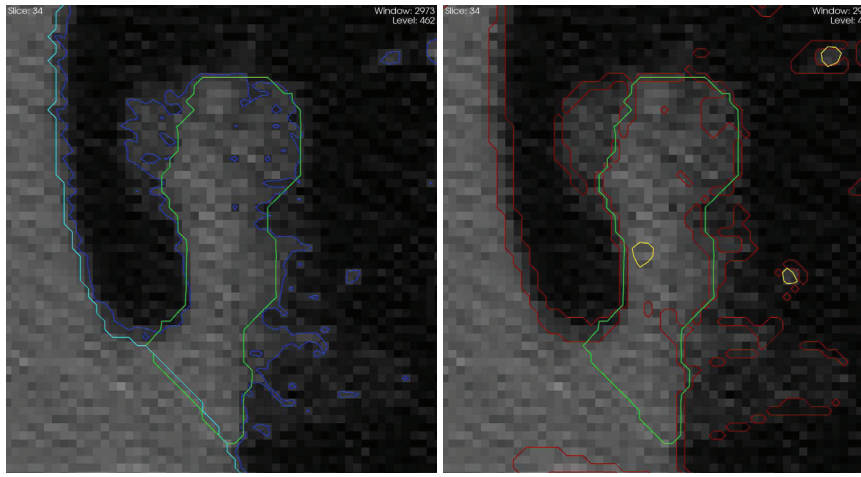


Fig. 4. Features (a) Lung wall feature in cyan, Intensity features in blue. (b)Vesselness features in yellow, Voxelized canny edge features in red, final segmentation in green. Note that the the final segmentation runs through the center of the voxelized canny edges as a result of cutting the level set at -0.5.

3.3 Zero change CT scans

Validation on clinical data was done on 6 coffee-break CT scans. (Table 1) These are CT scans of a patient obtained hours apart. (hence the name coffee-break). Ideally, segmentation would yield no change between the timepoints. However breathing artifacts, subtle differences in patient positioning, and noise result in slightly different segmentations. In addition, the two timepoints often had different acquisition parameters or resolution. The average ξ for the coffe break data was 10.9%.

Table 1. Zero Change Volumes V_1, V_2 for the 6 coffee break cases at two time points.

Case	$V_1 mm^3$	$V_2 mm^3$	Resolution (t1)	Resolution (t2)	$\xi = 200 \times \frac{V_1 - V_2}{V_1 + V_2}$
ST0108	1992.76	1990.41	$.703 \times .703 \times 1.25$	$.703 \times .703 \times 1.25$	0.11%
ST0109	2390.58	2698.9	$.703 \times .703 \times 1.25$	$.703 \times .703 \times 1.25$	12.11%
ST0111	3496.34	2484.37	$.703 \times .703 \times 1.25$	$.703 \times .703 \times 1.25$	33.84%
ST0112	290.07	297.62	$.703 \times .703 \times 1.25$	$.703 \times .703 \times 1.25$	2.56%
ST0113	670.27	654.42	$.703 \times .703 \times 1.25$	$.78 \times .78 \times 1.25$	2.39%
ST0114	3317.62	2864.81	$.56 \times .56 \times 1.25$	$.56 \times .56 \times 5$	14.64%

4 Toolkit and User Interface

The toolkit [13] is provided under a BSD license. It is cross platform and based on the Insight [14] and Visualization Tookits [15]. Fig. 5 shows the general architecture of the toolkit. Several feature generators and segmentation modules are provided, with an applicability towards a broader class of segmentation applications. A dashboard tests the algorithms against the aforementioned database providing a testbed to evaluate segmentation algorithms.

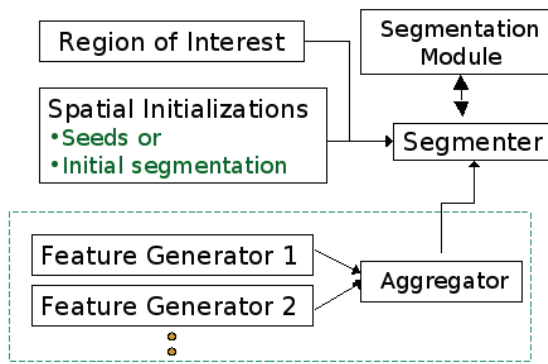


Fig. 5. Toolkit architecutre.

4.1 Front end user interface

VolView [16] is visualization tool from Kitware. It provides a front end to the tookit and features a guided wizard to perform one click segmentations of solid/part-solid lesions. The user identifies the region containing the lesion with a box, places a seed within the lesion and selects the lesion type (solid/part-solid). The application then segments the lesion and reports volumetrics. (Fig. 6.) A typical segmentation takes 5 to 25 seconds on a Intel 2.66GHz Core2Duo, 2GB RAM.

5 Results on the VOLCANO Grand Challenge data

For the VOLCANO Grand Challenge, all parameters remain as described above. No additional training was used to tailor the algorithm beyond the phantom data described in Section 3. User interaction consisted of the seed points and an ROI synthesized as a box $50mm$ on each side centered at the seed pixel.

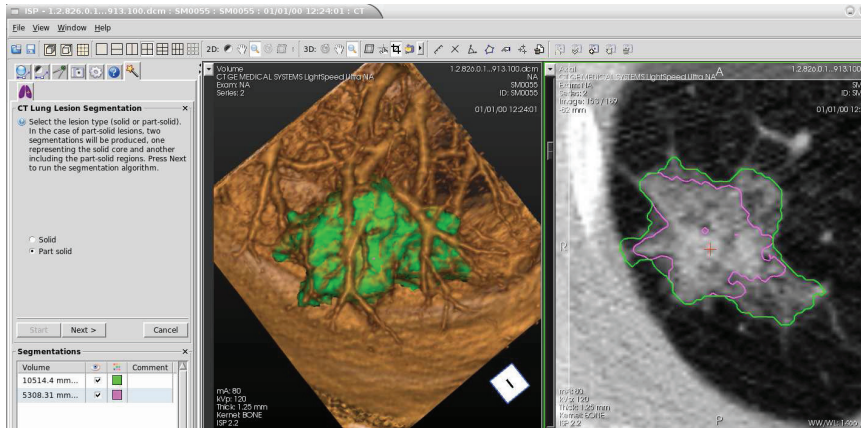


Fig. 6. A screenshot of the lesion sizing user-interface in VolView, on a part-solid lesion. The solid core (pink) and part-solid fringes (green) are segmented.

5.1 Generating the results

We provide the ability to automatically build and run the application against the Grand Challenge data. The user is expected to download the datasets from the VOLCANO [4] website. Instructions on obtaining the toolkit and building it may be found at [13]. Configure with the TEST_VOLCANO_DATA_COLLECTION option turned on. Build and run the tests. This will produce the volumetric estimates and segmentations for each of the datasets in the evaluation database.

5.2 Results

The results on the evaluation dataset are summarized in Table 2. Change is evaluated as $(V2 - V2)/V1$.

6 Conclusions

We have presented an algorithm to segment lung lesions from CT. We validated the results on clinical data and phantoms that provide a realistic anatomy. An important component of this work is reproducibility through open source, sharing of data and dashboards. Future directions may incorporate the use of scanner parameters. We will also validate against the spiculated and lobulated lesions in the anthropomorphic phantom, once ground truths are available.

7 Acknowledgements

We thank the Optical Society of America and the Air Force Research Laboratory for funding this work. We also thank David Yankelevitz (Weill Cornell Medical

Center) and Anthony Reeves (Cornell University) for the clinical and zero change cases. We also thank Nick Petrick and Lisa Kinnard, Marios Gavrielides and Kyle Myers from the FDA. The lung phantom data was provided by the U.S. Food and Drug Administration CDRH/OSEL/DIAM. The FDA collection was partially supported with funding through the intramural programs of NCI and NIBIB.from the FDA for the phantoms.

References

1. Therasse, P., et al: New guidelines to evaluate the response to treatment in solid tumors. J. National Cancer Institute **92**(3) (Feb 2000) 205–16
2. Erasmus, J.J., Gladish, G.W., Broemeling, L., Sabloff, B.S., Truong, M.T., Herbst, R.S., Munden, R.F.: Interobserver and intraobserver variability in measurement of nonsmall-cell carcinoma lung lesions: Implications for assessment of tumor response. J. Clinical Oncology **21** (Jul 2003) 2574–82
3. : WHO handbook for reporting results of cancer treatment. (1979)
4. : http://www.via.cornell.edu/challenge/.
5. Sato, Y., et al: 3D Multi-Scale line filter for segmentation and visualization of curvilinear structures in medical images. CVRMed MRCAS (1997) 213–22
6. Canny, J., F.: A computational approach to edge detection. IEEE Trans. Pattern Analysis and Machine Intelligence **8**(6) (1986) 679–698
7. Sethian, J.A.: Level Set Methods and Fast Marching Methods. Cambridge Press (1999)
8. Malladi, R., Sethian, J., Vemuri, B.C.: Shape modeling with front propagation: A level set approach. IEEE Trans. Pattern Anal. Machine Intel. **17**(2) (1995) 158–75
9. Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3d surface construction algorithm. In: SIGGRAPH. Volume 21. (July 1987) 163–169
10. Alyassin, A.M., Lancaster, J.L., Downs, J.H.: Evaluation of new algorisms for interactive measurement of surface area and volume. Medical Physics **21**(6) (1994)
11. Gavrielides, M.A., Zeng, R., Kinnard, L.M., Myers, K.J., Petrick, N.A.: A matched filter approach for the analysis of lung nodules in a volumetric CT phantom study. Proc. SPIE Med. Imaging (Feb 2009)
12. Gavrielides, M., Kinnard, L., Park, S., Kyprianou, I., Gallas, B., Badano, A., Petrick, N., Myers, K.: Quantitative in silico imaging and biomarker assessment using physical and computational phantoms: a review of new tools and methods available from the NIBIB/CDRH joint Laboratory for the Assessment of Medical Imaging Systems. Radiology (2008)
13. : Lesion sizing toolkit, http://public.kitware.com/LesionSizingKit.
14. Ibanez, L., Schroeder, W., Ng, L., Cates, J.: ITK Software Guide. Kitware Inc.
15. Schroeder, W., Martin, K., Lorensen, W.: Visualization Toolkit. Kitware Inc.
16. : Kitware Inc. Volview http://www.kitware.com/VolView.

Table 2. Results on the VOLCANO evaluation dataset

case	Change	V1 (mm^3)	V2 (mm^3)
SC0001	-0.93	4352.2	286.4
SC0002	-0.14	1942.94	1673.08
SC0003	-0.41	3805.55	2248.26
SC0004	0.55	3145.82	4873.41
SC0005	-0.12	2457.89	2161.53
SC0006	0.54	141.01	217.1
SC0007	0.01	393.48	398.88
SC0008	0.02	116.91	118.74
SC0009	-0.28	8029.24	5765.85
SC0010	-0.13	1187.27	1035.82
SC0011	0.06	1856.14	1964.2
SC0012	0.59	3016.81	4804.87
SC0013	0.04	16440.7	17177.9
SC0014	-0.11	8583.22	7635.9
SC0015	-0.17	400.99	331.64
SC0016	0.13	330.54	374.03
SC0017	1.05	1096.42	2252.93
SC0018	0.04	2289.68	2371.25
SC0019	0.54	2505.32	3861.29
SC0020	-0.02	1151.27	1125.71
SC0021	-0.01	1564.57	1551.21
SC0022	0.17	593.7	694.95
SC0023	-0.04	2172.98	2077.6
SC0024	0.14	820.46	938.44
SC0025	0.13	970.89	1094.68
SC0026	0.05	1493.97	1562.17
SC0027	0.13	12571.3	14195.3
SC0028	-0.06	1322.43	1244.15
SC0029	0.04	198.47	206.11
SC0030	-0.09	7628.96	6962.92
SC0031	-0.13	21902.6	19033.2
SC0032	-0.04	200.84	192.44
SC0033	0.1	481.6	528.31
SC0034	1.08	400.36	831.86
SC0035	-0.01	3021.52	2985.93
SC0036	0.08	892.11	966.66
SC0037	0.26	979.42	1236.19
SC0038	-0.02	1530.35	1506.32
SC0039	-0.33	1466.36	978.5
SC0040	-0.15	783.49	666.37
SC0041	-0.09	934.45	851.96
SC0042	0.12	584.12	652.95
SC0043	0.14	2338.12	2659.17
SC0044	1.46	438.45	1076.65
SC0045	0.03	678.82	697.13
SC0046	0.03	795.27	818.08
SC0047	0.13	1166.88	1317.97
SC0048	-0.25	511.91	384.87
SC0049	-0.35	4638.86	3004.46
SC0050	0.22	634.62	773.72

Automated approach to measure pulmonary nodule volume based on radius and CT number

Tatsuro Hayashi[1], Xiangrong Zhou[1], and Hiroshi Fujita[1]

[1] Department of Intelligent Image Information, Graduate School of Medicine,
Gifu University, 1-1 Yanagido, Gifu 501-1194, Japan
{hayashi, zxr, fujita} @fjt.info.gifu-u.ac.jp

Abstract. Determining the change in the pulmonary nodule size is a critical measurement for cancer diagnosis and therapy evaluation. In this study, an image-processing method that quantifies the nodule volume change based on computed tomography (CT) images is proposed. The proposed method consists of the following four steps: CT image interpolation, pulmonary region segmentation, nodule extraction, and nodule volume measurement. Nodule extraction is performed based on the radius and CT number. The proposed method was applied to a public database (Volcano '09). We confirmed that the nodule volume measurements performed using our proposed method were successful, according to a subjective evaluation.

Keywords: Pulmonary nodule, volume measurement, CT images, Volcano '09.

1 Introduction

Radiologists often use X-ray computed tomography (CT) images for detecting or evaluating pulmonary nodules. Determining the change in the pulmonary nodule size is a critical measurement for cancer diagnosis and therapy evaluation. One of the most important indicators of malignancy is the relative change in the size of a nodule over a period of time [1]. However, during nodule size estimation, there is a large amount of variation or disagreement in size estimation, even by medical experts. It seems that quantification of nodule volume using a computer would be effective in resolving this issue. In this study, a method that quantifies the nodule size change using relatively simple image-processing techniques is proposed.

2 Method

Figure 1 shows a flowchart of our proposed method. The method consists of the following four steps: (1) interpolation of the CT images, (2) segmentation of the pulmonary region, (3) extraction of the nodule region, and (4) quantification of the nodule volume.

The proposed method is a 3D-based approach that requires an isotropic image resolution in CT images. However, the spatial resolution of CT images often differs

between pixel spacing and slice thickness. To solve this problem, CT images are interpolated into an isotropic matrix by changing the slice thickness to pixel spacing as a pre-process of the nodule region determination. After this processing, the proposed method can be adaptable to any CT images with different spatial resolutions (Z-direction).

Pulmonary nodules exist within the pulmonary region. Therefore, the pulmonary region should be segmented first. The pulmonary region consists mostly of air, and it can be easily segmented using CT number thresholding. However, if the nodules are connected to the chest wall, determining the pulmonary region may not be easy. We designed a pulmonary segmentation method applicable to the attached nodules. The proposed method can detect various types of nodules (including attached nodules), and it is designed to adapt to the nodules with 2–30 mm diameters.

The proposed method is an automated approach and requires minimum user interaction (the user is only required to locate the center point of the target nodule). We identified nodules in the high-CT number regions on the pulmonary regions. Therefore, the proposed method extracts nodules using CT number thresholding. However, making a distinction between nodules and blood vessels using CT numbers is difficult. The proposed method attempts to solve this problem by localizing the nodule existence range before gray-level thresholding.

The details of the proposed method are described below.

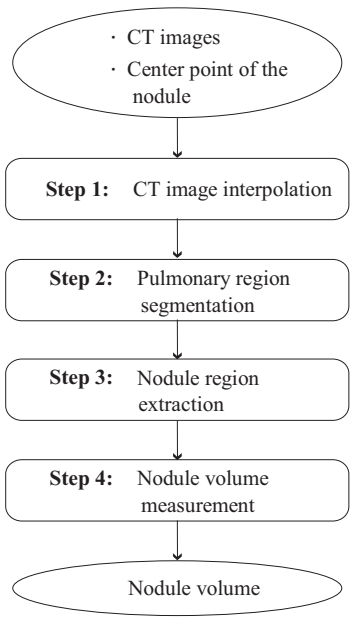


Fig. 1. Flowchart of the proposed method.

2.1 CT image interpolation

In this section, we describe an interpolation method by which each voxel on the CT images is interpolated to have isotropic dimensions. Let *s* be the spatial coordinates of the image, which is defined as follows:

$$s = \{(x, y, z) : 1 \le x \le w, 1 \le y \le h, 1 \le z \le d\}. \quad (1)$$

Let $f(p) \mid p \in s$ be an arbitrary image. First, the pixel spacing and slice thickness are obtained from a DICOM file header. Next, the CT images are reformatted using the following equation to conform the slice thickness to the pixel spacing.

$$f(x_0, y_0, z_0) = \sum_k \sum_l \sum_m f(x_k, y_l, z_m) C(x_k - x_0) C(y_l - y_0) C(z_m - z_0), \quad (2)$$

where (x_k, y_l, z_m) denotes the grid point around (x_0, y_0, z_0). The interpolating function *C* is defined as sinc function [2].

$$C(x) = \frac{\sin \pi x}{\pi x}. \quad (3)$$

Figure 2 shows an example of before and after interpolating the CT images. In this figure, a change in the slice thickness in the CT images can be recognized. In the following sections, we use *CT*(*p*) to represent the interpolated CT images.

(a) Before interpolation (b) after interpolation

Fig. 2. An example of interpolated CT images (one coronal section).

2.2 Pulmonary region segmentation

In this section, we describe a method for segmenting the pulmonary region (see Fig. 3). The pulmonary region consists mostly of air. Therefore, the pulmonary region can be identified as a region with low CT numbers on the CT images. Therefore, the proposed method extracts air regions using CT number thresholding. Threshold *th* H.U. is determined using the Otsu method [3]. However, the existence of noise on the CT images often causes segmentation failure. In order to reduce the noise, the CT images are smoothed using a median filter. A binary image *B*(*p*) is then generated using the following equation.

$$B(p) = \begin{cases} 1 & if\ Med(CT(p)) < th \\ 0 & otherwise \end{cases}, \quad (4)$$

where *Med* denotes the function of the median filter.

Each air region is then labeled by applying 3D labeling processing to *B*(*p*). Then, external air regions and small air regions are removed, and the pulmonary region *B'*(*p*) is extracted as follows:

$$B'(p)=\begin{cases}1 \;\; if \; A_i=0 \wedge \sum q \geq \alpha \mid q \in Label(B[p])=i, 1 \leq i \leq n \\ 0 \;\; otherwise\end{cases}, \tag{5}$$

$$A_i=\begin{cases}1 \;\; if \; Cir(q)=1 \mid q \in Label(B[p])=i, 1 \leq i \leq n \\ 0 \;\; otherwise\end{cases}, \tag{6}$$

$$Cir(p)=\begin{cases}1 \;\; if \left(x-\frac{w}{2}\right)^2+\left(y-\frac{h}{2}\right)^2=r^2 \\ 0 \;\; otherwise\end{cases}, \tag{7}$$

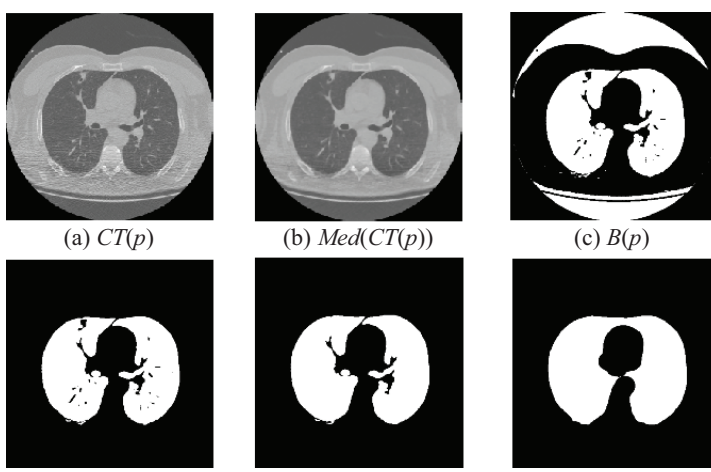
where r voxel denotes the radius of the circle.

Holes in the pulmonary region are then filled in each slice. In particular, each voxel value of $B'(p)$ is reversed, and labeling processing is performed. Labeled regions that are not adjacent to the edge of the slice image are added to the pulmonary region. This resulting image is defined as $B''(p)$.

Finally, closing operation [4] is applied to $B''(p)$, and the binary image of the pulmonary region $B_p(p)$ is obtained.

$$B_p(p)=\left[\left(B''(p)\oplus g^s\right)\Theta g\right](p), \tag{8}$$

where $\oplus$ denotes the Minkowski sum, Θ denotes the Minkowski difference, and g denotes the structural element, which consists of a sphere with a radius of t mm.



(a) $CT(p)$ (b) $Med(CT(p))$ (c) $B(p)$

(d) $B'(p)$ (e) $B''(p)$ (f) $B_p(p)$

Fig. 3. Process flow of pulmonary region segmentation (one axial section).

2.3 Nodule region extraction

In this section, we describe a extraction method of the nodule region (see Fig. 4). First, CT number thresholding is applied to the pulmonary region, and the image $N_1(p)$ is obtained.

$$N_1(p) = \begin{cases} 1 \ if \ Med(CT(p)) \geq th_2 \wedge B''(p) = 1 \\ 0 \ otherwise \end{cases}, \quad (9)$$

where th_2 H.U. is determined by applying the Otsu method to the smoothed CT voxels with $B''(p)=1$.

Second, Euclidean distance transformation [5] is applied to $N_1(p)$, and the distance image $D(p)$ is obtained.

$$D(p) = \min_{q=(x_0,y_0,z_0)\in s} \sqrt{(x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2} \mid N_1(q) = 0. \quad (10)$$

Blood vessels have a tubular-like structure, while nodules have a mass-like ("spherical") structure. Therefore, the distance value of the center of a nodule should be larger than the center of a blood vessel. Therefore, voxels within the distance value $D(c)$ are removed. Here, $c \in s$ is defined as the center point of the nodule that the user inputted. However, there is no guarantee that the user always inputs the center of the nodule. The adjusted center point $c' \in s$ is thus obtained using the following program.

```
program Optimizing center point(Input: c, Output: c')
  {c, c', c1, and c2∈s,
  neighbor(p): The function which outputs the adjacent
  coordinates of the point p};
  var   c1, c2;
  begin
    c2 = c;
    repeat
      c1 = c2;
      if(D(c2) < D(neighbor(c1)))
        c2 = neighbor(c1);
      )
    until c1 = c2
    c' = c1;
  end.
```

Third, $D'(p)$ is obtained by the following equation.

$$D'(p) = \begin{cases} D(p) \ if \ D(p) \geq D(c') \\ 0 \quad otherwise \end{cases}. \quad (11)$$

Fourth, $D'(p)$ is applied to reverse Euclidean distance transformation [6], and image $N_2(p)$ is obtained.

$$N_2(p)=\begin{cases}1 \ if \ d(p,q)<D'(q) \mid q\in s \\ 0 \ otherwise\end{cases}, \tag{12}$$

where $d(p, q)$ is the function that outputs the Euclidean distance between points p and q.

Fifth, $N_2(p)$ is applied to labeling processing, and only the region including c' is extracted. This resulting image is defined as $N_3(p)$.

$$N_3(p)=\begin{cases}1 \ if \ Label(N_2(p))=Label(N_2(c')) \\ 0 \ otherwise\end{cases}. \tag{13}$$

If a nodule is a pure sphere, a nodule region will be extracted in $N_3(p)$. However, since a nodule is not a pure sphere, it is necessary to extend the region. At the sixth processing, the image $N_4(p)$ that extended the region is obtained.

$$N_4(p)=\begin{cases}1 \ if \ d(p,q)<D''(q) \mid q\in s \\ 0 \ otherwise\end{cases}, \tag{14}$$

$$D''(p)=\begin{cases}D(p) \ if \ N_3(p)=1 \\ 0 \quad otherwise\end{cases}. \tag{15}$$

Seventh, the $N_4(p)$ region is applied to CT number thresholding, and the nodule region $N_5(p)$ is obtained.

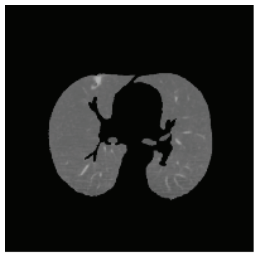
$$N_5(p)=\begin{cases}1 \ if \ N_4(p)=1\wedge CT(p)\geq th_3 \wedge B_p(p)=1 \\ 0 \ otherwise\end{cases}, \tag{16}$$

where th_3 H.U. is the fixed threshold.

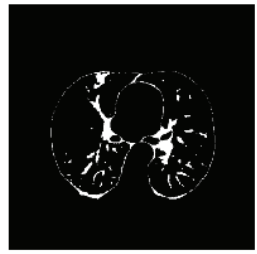
The holes in $N_5(p)$ are then filled using the abovementioned method. The resulting image is defined as $N_6(p)$. Finally, the opening operator [4] is applied to $N_6(p)$, and then, only the region including c' is extracted. Hence, the nodule region $N(p)$ is obtained.

$$N_7(p)=\left[\left(N_6(p)\Theta g_2^s\right)\oplus g_2\right](p), \tag{17}$$

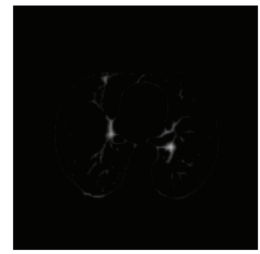
$$N(p)=\begin{cases}1 \ if \ Label(N_7(p)=Label(N_7(c'))) \\ 0 \ otherwise\end{cases}. \tag{18}$$



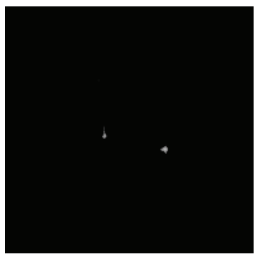
(a) $Med(CT(p)) \cap B''(p) = 1$



(b) $N_1(p)$

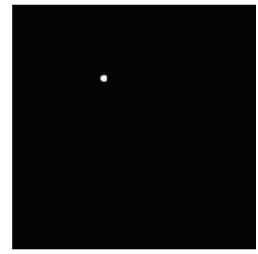


(c) $D(p)$

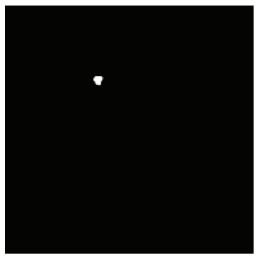


(d) $D'(p)$

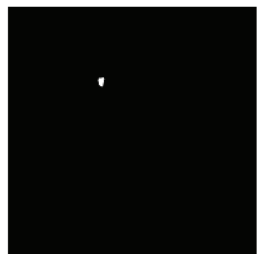
(e) $N_2(p)$



(f) $N_3(p)$



(g) $N_4(p)$



(h) $N(p)$

Fig. 4. Process flow of nodule region extraction (one axial section).

2.4 Nodule volume measurement

The nodule volume V is calculated using $N(p)$ and pixel spacing.

$$V = \sum q \times pixel\_spacing^3 \mid q \in s \wedge N(q) = 1. \tag{19}$$

3 Results

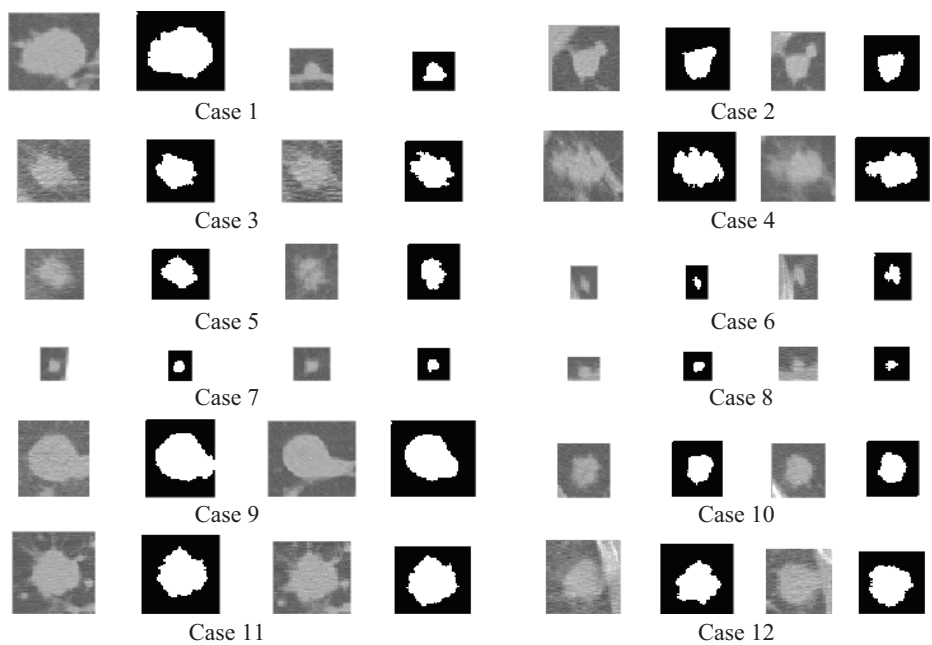
The proposed method was applied to a database known as Volcano '09. Volcano '09 consists of CT images obtained from 50 cases. Each CT case was scanned over a

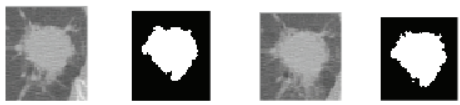
period of time. Let V_1 and V_2 be the volume of the nodules in the CT images for the first and second scans, respectively. Note that the target nodule is defined in advance. In order to quantify the nodule volume and its change, we calculated V_1, V_2, and $(V_2 - V_1)/V_1$ using the proposed method. Parameters used in the experiment were $r = 255$, $t = 15$, and $th_3 = -350$. Moreover, the smoothed image was defined as the CT image that was applied by using a median filter with a mask size of $3 \times 3 \times 3$, twice. These parameters were determined empirically.

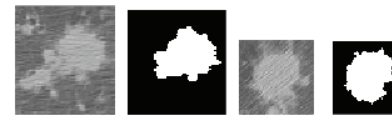
The results of the extracted nodules are shown in Fig. 5, and the nodule volume measurement results are shown in Table 1. By carrying out a subjective evaluation, we confirmed that the nodule volume measurements were successful in most cases. However, cases in which the nodule size was small produced a large measurement error due to over- or under-extraction. For example, the outside of the pulmonary region was slightly over-extracted in the second scan in case 6. Moreover, the size of this nodule was quite small. Therefore, the value of an excessive size ratio was calculated.

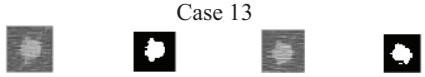
It was difficult to determine the contour of the nodules in cases 30 and 31 because those nodules arose in the pulmonary hilum. Values of the CT numbers of nodules and tissues in the pulmonary hilum are almost the same; therefore, separating the nodule from the pulmonary hilum based on the CT numbers is difficult. However, the anatomical shapes of the pulmonary hilum could be used to solve this problem.

The calculation time in each case was 59 ± 21 s (CPU: Intel Core2 Duo E7200, Memory: 4 GB). A majority of the calculation cost (approximately 90%) was spent in the smoothing process. If an alternative smoothing method with low calculation cost is used instead of a method that involves the use of a median filter, the calculation cost will be reduced.




Case 13


Case 14


Case 15


Case 16


Case 17


Case 18


Case 19


Case 20


Case 21

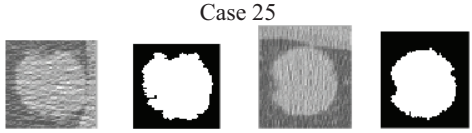

Case 22


Case 23


Case 24

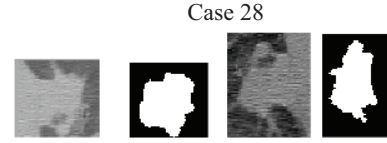

Case 25

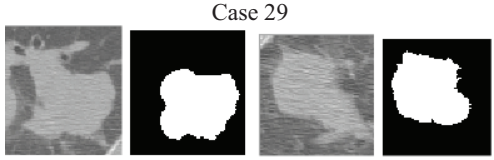

Case 26


Case 27


Case 28


Case 29

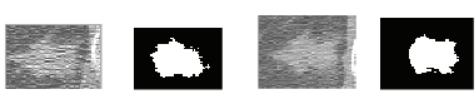

Case 30


Case 31


Case 32


Case 33


Case 34


Case 35


Case 36


Case 37


Case 38

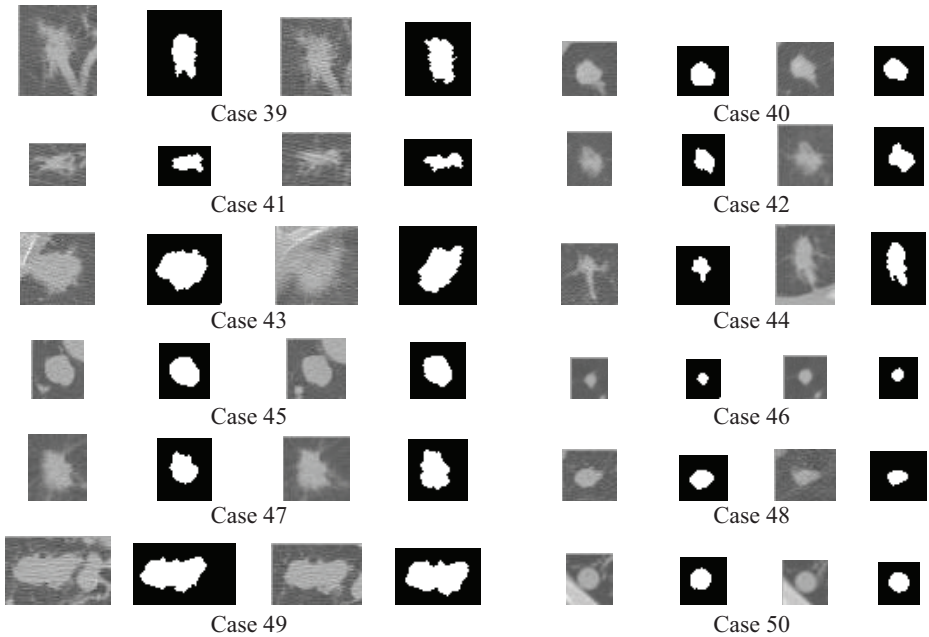


Fig. 5. Nodule extraction results for all 50 cases using our proposed method (one axial section around a target nodule). From left to right, the figure shows the interpolated CT images in the first scan, the extraction results of a target nodule in the first scan, the interpolated CT images in the second scan, and the extraction results of a target nodule in the second scan.

4 Discussions

In the case of lesion size or volume estimation, there is a large amount of variation and disagreement about the results, as determined by different experts. We believe that quantitative size or volume estimation carried out using a computerized method can contribute to reduce such variation and disagreement. A computerized method with a few user interactions is desirable, because such interactions are dependent on the user's subjectivity. Therefore, we designed the proposed method that requires minimum user interaction (the user must only indicate a point near to the center of target nodule). Furthermore, the proposed method can automatically adjust the point from the position of user interventions to the "real center" of the target nodule. The position of the points from user interventions had observer variations, and the proposed method could provide results independent of those variations.

We only selected and used one feature for nodule segmentation in the proposed method, based on the experience of radiologists. That is, single CT number thresholding was used to extract high-CT number regions, which radiologists would interpret as nodule regions on CT images. However, it was difficult to discriminate the pulmonary nodules from other structures using CT number thresholding in some

Table 1. Measurement results of the volumes V1 and V2 together with the volume ratio.

CaseID	(V2 − V1)/V1	V1 (mm^3)	V2 (mm^3)
ST0001	−0.96	3642	161
ST0002	−0.19	1431	1153
ST0003	−0.21	1842	1461
ST0004	0.09	2155	2360
ST0005	0.11	949	1049
ST0006	2.90	26	103
ST0007	−0.44	97	54
ST0008	−0.04	36	34
ST0009	−0.30	5720	3983
ST0010	−0.02	635	622
ST0011	0.08	1452	1565
ST0012	1.06	1770	3652
ST0013	−0.02	8121	7972
ST0014	−0.18	6036	4968
ST0015	−0.33	214	142
ST0016	−0.14	170	147
ST0017	1.60	452	1175
ST0018	−0.01	1978	1949
ST0019	0.13	1610	1813
ST0020	−0.04	824	792
ST0021	−0.00	1158	1157
ST0022	0.22	284	347
ST0023	−0.05	1771	1680
ST0024	0.24	562	700
ST0025	0.18	525	620
ST0026	0.31	766	1007
ST0027	−0.09	12842	11723
ST0028	−0.33	656	443
ST0029	−0.03	26	25
ST0030	−0.16	6407	5370
ST0031	−0.12	16381	14394
ST0032	0.21	27	33
ST0033	0.06	337	357
ST0034	−0.42	86	50
ST0035	−0.01	1951	1939
ST0036	0.00	723	724
ST0037	0.17	688	802
ST0038	0.04	1157	1202
ST0039	−0.19	718	582
ST0040	−0.17	498	413
ST0041	0.08	458	493
ST0042	0.46	200	292
ST0043	0.09	1794	1953
ST0044	2.31	207	684
ST0045	0.07	457	490
ST0046	0.06	59	62
ST0047	0.40	504	704
ST0048	−0.47	317	168
ST0049	−0.08	2578	2372
ST0050	−0.10	953	857

attached nodules or nodules with blood vessels. Therefore, the proposed method attempted to determine the nodules' boundaries based on their shape (radius). This approach produced appropriate results in most of the cases, as shown in Fig. 5. However, there were some cases with over-extraction results (e.g., the first scan in case 47) or under-extraction results (e.g., in cases 30 and 31). Designing a complicated method, such as a specialized scheme for nodule shapes or types of circumference tissues, may be required in order to improve the segmentation accuracy of the nodule.

Some features are assumed to be effective in nodule size estimation, including radius, major/minor axis, the nodule area in its center slice, and so on. This research used the voxel number of the 3D volume of nodules to measure the change in the nodule size because the 3D volume can accurately reflect the shape of the different nodules. The other features derived from the 3D volume, such as the degree of sphericity, should also be effective for nodule diagnosis and therapy evaluation.

5 Summary

In this study, a method that quantifies the nodule volume change by using relatively simple image-processing techniques was proposed. The nodule extraction was performed based on radius and CT number. The proposed method was applied to a database known as Volcano '09. By carrying out a subjective evaluation at an initial stage, we confirmed that the nodule volume measurements were successful.

Acknowledgments

The authors thank members of the Fujita Laboratory for their valuable discussions, especially Dr. Fukuoka. This research was supported in part by a research grant from Kayamori Foundation of Informational Science Advancement, in part by a research grant from Gifu University, and in part by a research grant of Grant-in-Aid for Young Scientists B (21700462) from Japan Society for the Promotion of Science (JSPS).

References

1. VOLCANO'09 Challenge, http://www.via.cornell.edu/challenge/index.html
2. Tamura, H.: Computer Image Processing, 127, Ohmsha (2002) (in Japanese)
3. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Systems Man Cybernet, 9 (1), 62-69 (1979)
4. Kobatake, H.: Morphology, Corona (1996) (in Japanese)
5. Saito, T., Toriwaki, J.: Euclidean distance transformation for three dimensional digital images. IEICE D-II, J76-D-II (3) 445-453 (1993) (in Japanese)
6. Saito, T., Toriwaki, J.: Reverse Euclidean distance transformation and extraction of skeletons in the digital plane. IEICE Technical Report, Pattern Recognition and Understanding, 93 (228): 57-64 (1993) (in Japanese)

Nodule volume change estimation in thoracic CT using sphere fitting, morphological segmentation and image registration

Thomas Duindam[1], Bartjan de Hoop[2], and Bram van Ginneken[1]

[1] Image Sciences Institute, University Medical Center Utrecht, the Netherlands
[2] Department of Radiology, University Medical Center Utrecht, the Netherlands

Abstract. Three methods are presented to assess the relative volume change of a pulmonary nodule between two chest CT scans when the nodule location in both scans is provided. The first method fits a sphere around both nodules and computes the volume of dense tissue in that sphere. The second method segments both nodules using thresholding, component labeling and morphological processing. The third method applies non-rigid registration to transform the first to the second scan and applies that transformation to a segmentation of the nodule in the first scan to obtain a segmentation of the nodule in the second scan. All methods are applied to 50 nodule pairs from the VOLCANO'09 challenge. These cases are divided by a radiologist in stable and growing pairs. All methods produce lower mean volume change for the stable cases compared to the growing nodules, but the distributions overlap considerably. Moreover, the correlation between the volume change estimates produced by the three methods is modest. This shows that nodule volume change assessment is a complicated problem.

1 Introduction

Pulmonary nodules occur frequently in thoracic CT scans. In studies where subjects at high risk for developing lung cancer were scanned with low-dose CT, 8% to 51% of all subjects had at least one nodule [1]. The most important question that follows detection of a nodule is whether the lesion may be malignant. Comparison with available prior imaging of the same nodule is a critical step in answering this question. Malignant lesions tend to have a volume doubling time between 20 and 400 days. Nodules with more rapid growth likely represent inflammation and nodules that are stable for a long period of time are likely benign. Nodule growth rate is one of the most important characteristics in the determination of its probability of malignancy. Therefore, a substantial amount of research has been devoted to the design of methods to accurately estimate the growth rate of nodules. In this paper, growth ΔV is defined as the proportional change in size of the lesion between the two scans relative to the size of the lesion in the first scan. Let V_1 and V_2 be the volume of the lesion in scan 1 and 2, respectively, then we define the growth as

$$\Delta V = \frac{V_2 - V_1}{V_1}. \quad (1)$$

In this definition ΔV can take values from -1 (the nodule vanishes completely) to 0 (nodule is stable) to 1 (nodule doubles in volume) to infinity. Because the interval [-1,0] corresponds to $[0,\infty]$, values for ΔV are often not normally distributed.

The higher the accuracy of the growth estimation method, the shorter the time window can be between a baseline scan where a nodule is identified and a follow-up scan used to determine growth rate. For example, if a nodule has a doubling time of 100 days, its ΔV will be around 0.22 after 30 days. If the error in ΔV is also around 0.2, one cannot conclude decisively after only one month whether a measured growth is real or a measurement error. Waiting for three months would be enough in this example scenario, but if the nodule was a fast growing cancer with a doubling time of 20 days, a three month delay in detection may be fatal. Most ongoing trials for lung cancer screening use two to six months follow-up scans in case an intermediately sized nodule (with a volume around 50 to 500 mm^3) has been detected. If more accurate methods for growth assessment were available, this interval could be shortened.

The data used in this paper is from the Volcano '09 Challenge (`http://www.via.cornell.edu/challenge/`). The Volcano website states the goal of the challenge as follows:

> The goal of this study is to compare the outcomes of various algorithms measuring the change in volume of pulmonary nodules from CT scans using a common dataset and performance evaluation method.

The dataset used contains two CT-scans of each case, taken at different, unknown points in time. The location of the approximate center of the nodule is provided in an accompanying file. Image resolution, nodule size and scan parameters vary from case to case.

An obvious approach for calculating the change of volume in a nodule over time is to segment the nodule in both scans (or in every scan if more than two scans are available) and compare the computed volumes. Nodule segmentation is not a trivial task however. Several studies have shown that when the same nodule is segmented in two scans taken only minutes apart, ΔV can vary substantially. De Hoop et al. [2] have shown in an evaluation of six commercially available lung nodule segmentation toolkits that a difference of 20% is required to be able to say confidently that a nodule has truly changed (the 20% corresponded to twice the standard deviation of the ΔV measurements).

One important reason why it is difficult to automatically segment nodules is that other dense structures are often attached to the nodule. These structures, which can be vessels, the pleural surface, fissures or dense abnormalities, could be erroneously included in the nodule segmentation. If this would happen in both segmentations the estimated volume change might still be correct. But it is possible that a nodule in a baseline scan is not in direct contact with a

vessel, but in a follow-up scan due to nodule growth or differences in inspiration level, the nodule is in contact with a vessel and that vessel is included in the segmentation. Such a segmentation inconsistency will cause the extent of nodule growth to be overestimated. This has been pointed out by Reeves et al. [3] and they have proposed a method to repair such inconsistencies. In this work we compare a nodule segmentation method (SEGM) which independently segments the nodule in both scans to compute growth with two other approaches which may be less sensitive to segmentation inconsistencies.

Another approach is to define an equally sized volume of interest in both scans and apply a threshold to the voxels inside the volume. If we assume that the volume of the vasculature surrounding the nodule remains stable over time, an estimation of nodule volume change can be made by counting the voxels remaining in the volume of interest after applying the threshold. In this work, we use a sphere for the volume of interest. We shall refer to this method as SPHERE.

One other approach is to segment the nodule only in the first scan and use the deformation field obtained by non rigidly registering both scans to transform the first segmentation. In this way, a segmentation for the second nodule is derived. This method has proven to be successful for artificial nodules [4]. The strength of this approach is that slight errors the segmentation will not lead to large errors in the measured relative volume change, as long as the nodules can be registered successfully. For example, if a part of an attached vessel is erroneously included in the first segmentation, the transformed segmentation will also include that vessel part and it will probably not be changed much. This method will be referred to as REG throughout this paper.

We present an evaluation of the three aforementioned approaches to nodule volumetry. The output of all three methods with various parameter settings has been evaluated, visually assessed, and compared with visual evaluation of nodule change by a human expert. Section 2 provides some details about the data used in the VOLCANO challange. The methods are detailed in Section 3. Experiments and results are presented in Section 4. We discuss the results and draw conclusions in Section 5.

2 Data

The evaluation dataset consists of 50 pairs of nodules. Many pairs have been taken during the initial stages of a lung biopsy and should therefore not exhibit any growth. For others it is known, through follow-up, that they have grown. Data originated from Weill Cornell Medical College, New York, USA. Data is divided into three categories. The first category consists of 27 nodules visible on two scans of 1.25 mm slice thickness with little observed size change, and a range in diameter from approximately 4 to 24 mm. The second category of nodules included 13 nodules imaged on either two 2.5 mm scans or one 1.25 mm scan and one 2.5 or 5.0 mm scan to examine the effect of slice thickness on the performance. The nodules ranged in size from approximately 8 to 30 mm. The

third category consists of an additional 9 nodules on two 1.25 mm scans, but with a large size change; these nodules ranged in size from approximately 5 to 14 mm.

Only a limited number of slices were made available, and a list of locations of the nodules in both scans. It was not given to which of the three categories above the nodule pair belonged. Data was anonymized, and the scan order (which scan was made first in time) was not given. Our results were made available to the VOLCANO organizers for further analysis.

To be able to analyze the results of our methods, we asked a radiologist to visually inspect the cases. He used a side-by-side viewer that was developed in our group, which also allowed to watch the scan pairs after rigid registration. He assigned each pair to one of the follwing five classes: definitely shrinkage, possibly shrinkage, stable, possibly growth and definitely growth. For our subsequent analysis we grouped the cases of possible and definite growth/shrinkage, and swapped all pairs with shrinkage so that two groups remained: stable nodules (28 cases) and growing nodules (22 cases).

3 Method

All methods employ the result of a lung segmentation. Typically a 3D lung segmentation algorithm would be used for this purpose. However, the scans in the VOLCANO challenge only cover a limited axial field of view. Therefore a 2D lung segmentation was applied slice by slice. The segmentation method consists of a sequence of standard image processing steps (thresholding, component labeling, hole filling, morphological closing) and is described in [5].

3.1 Morphological Segmentation (SEGM)

If a segmentation of the nodule is in both scans is present, computing the change in volume is a trivial task. The segmentation algorithm used for this study is a similar and somewhat simplified version of the algorithm described in [6].

The first step of the segmentation algorithm is very similar to the SPHERE approach: first a $50\times50\times50$ mm volume of interest is super sampled to isotropic voxels of 0.5 mm, followed by a thresholding operation with threshold t using the lung segmentation as a mask. The next steps of the algorithm aim to remove vasculature and noise from the nodule segmentation. First a connected component analysis is executed to isolate the largest connected component. This removes noise and disconnected vessels from the segmentation. After this step only the nodule and vessels with connections to the nodule remain. To remove these vessels mathematical morphology is used. With an opening with a spherical kernel, vessels connected to the nodule are removed. The diameter d of this kernel is the second parameter for the segmentation algorithm. The opening will also remove voxels on the edge of the nodule, and smooth its surface. To reduce this effect a conditional dilation is used with an decreasing kernel size [6]. This will regrow the nodule back to its original size and reconstruct most features of the nodule

surface. Not all features can be reconstructed, but since relative volume change is more relevant than the absolute volume of a nodule this is deemed acceptable. It is possible that the opening will completely remove all candidate voxels from the segmentation, in this case all morphological operations are canceled, and the result of the connected component analysis is used for further computations. To ensure segmentation consistency as much as possible, we kept the segmentation parameters t and d the same for both scans in a pair.

3.2 Sphere fitting (SPHERE)

The underlying assumption of this technique is that apart from the nodule, the volume of anatomical structures inside the lung should not have changed significantly over time. Although it seems unlikely that this assumption will hold over the entire lung, especially in patients with gross pathology, there will likely be a small volume of interest around the nodule for which this assumption will hold. If we can find such a volume of interest for both scans, simply counting the number of dense voxels will suffice for making an estimation of nodule growth.

The first step of this algorithm is to take a $50\times50\times50$ mm volume of interest centered at the nodule center, which is provided in the Volcano data set. This volume of interest is super sampled to isotropic voxels of 0.5 mm resolution. This reduces partial volume effects and also simplifies calculations as the voxels in both scans now have equal dimensions.

Next, a threshold t is applied over this volume of interest, excluding voxels outside the lung field, where we use the predetermined lung segmentation as a mask. This leaves us with all voxels denser than t inside the lung (we call these dense lung voxels). Effectively this means that only nodule and vessel voxels remain. If we would compute the volume change at this point we would not be able to quantify growth accurately, since the volume of the nodule might be very small compared to that of the surrounding vasculature. Therefore we refine the volume of interest as follows. We define a spherical volume of interest at the approximate nodule-center, and iteratively increase the diameter of this sphere by two voxels (1, 3, 5, and so on). We keep increasing the diameter until the next iteration would not add a sufficient number of dense lung voxels. We define this cut-off ratio r as the number of dense lung voxels added in the next iteration divided by the total number of voxels added in the next iteration. Thus this algorithm has two parameters, the threshold t and the cut-off ratio r.

This leaves us with two spherical volumes of interest, possibly with different diameters. This will introduce a bias in the measurement. The larger of the two is therefore chosen and applied to both scans. The final volume in mm^3 is than computed by multiplying the number of dense lung voxels in the spheres by 0.5^3.

3.3 Segmentation by registration (REG)

In this approach we use a segmentation of the nodule in one scan (the moving scan), and to avoid segmentation inconsistencies that could occur if the same algorithm were applied independently to the other (fixed) scan, we instead use

image registration to elastically deform the moving so as to resemble the fixed scan. We use the freely available package elastiX [7], version 3.9 for registration. The registration software is able to handle anisotropic voxel sizes internally.

To force the registration software to align only nodules and not the surrounding tissue a mask is used. To make the mask, first both nodules are segmented using the morphological segmentation algorithm outlined above. This resulting segmentation is then dilated to ensure that the entire nodule is represented in the mask, even if under-segmentation has occurred. To ensure that the same structures are present in both masked areas the biggest of the two masks is used for both scans. Finally the parts of the mask that are outside the lungs are removed using the lung segmentation mentioned before.

The resulting transformation is applied to the segmentation of the nodule in the moving scan, and both volumes are compared. This method is also described in [4], and has proven to be effective for artificial nodules. It is not clear what mask, if any, was used for the experiments in [4]. The only parameters the REG method takes are the threshold t and the kernel size d for the segmentation.

Setting for the registration method were normalized mutual information as similarity measure, first a rough alignment with an affine transformation, followed by a non-rigid registration modeled by B-splines. Three resolutions were used in both affine and non-rigid stages, and always 500 iterations were performed per resolution level. The other settings were the defaults of the package.

4 Experiments and results

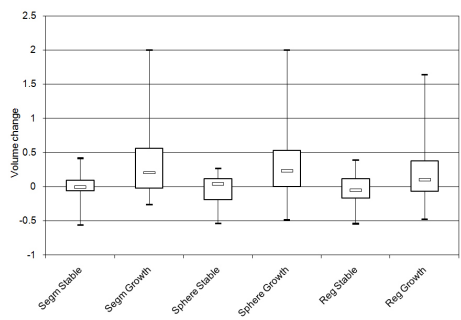


Fig. 1. ΔV of all three methods for stable and growing nodules. Here parameters were determined by a user. Whiskers indicate minimum and maximum values.

Two types of experiments have been carried out. First, for all three methods, several parameter settings were evaluated. Table 1, 2 and 3 give the results.

Second, all three methods were applied using an optimal setting for the two parameters of each method, determined visually by a human observer. This

Table 1. Variability of ΔV outcome for the SPHERE method using different values for threshold t and cutoff r. Stable and growth cases are treated separately. For each setting the average, standard deviation(SD), minimum, 1st quartile, median, 3rd quartile and the maximum value are reported. The row with label semi auto are the results from the semi-automatic experiments.

t	r	**Avg**	**SD**	**Min**	**q1**	**Median**	**q3**	**Max**
		Stable nodules						
-600	0.5	0.118	0.389	-0.377	-0.031	0.015	0.199	1.263
-600	0.6	0.057	0.378	-0.550	-0.111	0.010	0.111	1.263
-600	0.9	0.181	0.602	-0.618	-0.087	-0.004	0.289	2.000
-500	0.5	0.015	0.261	-0.422	-0.065	0.005	0.071	0.792
-500	0.6	0.003	0.231	-0.573	-0.060	0.018	0.117	0.445
-500	0.9	0.168	0.621	-0.378	-0.116	-0.004	0.094	2.000
-400	0.5	-0.010	0.231	-0.619	-0.090	0.005	0.061	0.423
-400	0.6	-0.007	0.253	-0.668	-0.085	0.021	0.101	0.446
-400	0.9	-0.011	0.326	-0.465	-0.234	-0.029	0.082	0.928
Semi	auto	-0.024	0.217	-0.540	-0.190	0.041	0.118	0.271
		Growing nodules						
-600	0.5	0.001	0.179	-0.433	-0.148	0.012	0.134	0.257
-600	0.6	0.000	0.205	-0.507	-0.174	0.030	0.137	0.462
-600	0.9	0.053	0.468	-0.507	-0.091	-0.002	0.039	2.000
-500	0.5	-0.015	0.201	-0.539	-0.171	0.035	0.112	0.239
-500	0.6	0.020	0.331	-0.539	-0.179	0.039	0.100	1.239
-500	0.9	0.066	0.493	-0.609	-0.107	0.016	0.035	2.000
-400	0.5	-0.023	0.224	-0.572	-0.193	0.042	0.116	0.271
-400	0.6	0.021	0.360	-0.572	-0.193	0.046	0.124	1.319
-400	0.9	0.094	0.550	-0.650	-0.163	0.014	0.043	2.000
Semi	auto	0.381	0.634	-0.485	0.003	0.232	0.532	2.000

makes our approaches semi-automatic, but the amount of interaction required is minimal. Figure 1 shows the distribution of ΔV for growing and stable nodules for all methods. Note how the REG method works well for stable nodules (i.e. ΔV close to zero), but not for growing nodules, where q1 is below zero. The results for the SEGM and the SPHERE method are similar in the case of growing nodules, but for stable nodules the distribution for the SEGM method has a lower variability.

Figures 2 to 4 provide examples of the results of each method, including a case where the SEGM method leads to segmentation inconsistencies. Scatterplots of the results of different methods for stable and growing cases are given in Figures 5 and 6. Correlation between SEGM and REG is 0.667, between REG and SPHERE 0.479 and the highest correlation is achived between SEGM and SPHERE with 0.832.

Table 2. Variability of ΔV outcome for the SEGM method using different values for threshold t and kernel size d. Stable and growth cases are treated separately. For each setting the average, standard deviation(SD), minimum, 1st quartile, median, 3rd quartile and the maximum value are reported. The row with label semi auto are the results from the semi-automatic experiments.

t	d	**Avg**	**SD**	**Min**	**q1**	**Median**	**q3**	**Max**
-600	3	-0.118	0.389	-1.263	-0.199	-0.015	0.031	0.377
-600	4	-0.057	0.378	-1.263	-0.111	-0.010	0.111	0.550
-600	6	-0.181	0.602	-2.000	-0.289	0.004	0.087	0.618
-500	3	-0.015	0.261	-0.792	-0.071	-0.005	0.065	0.422
-500	4	-0.003	0.231	-0.445	-0.117	-0.018	0.060	0.573
-500	6	-0.168	0.621	-2.000	-0.094	0.004	0.116	0.378
-400	3	-0.122	0.837	-2.000	-0.255	-0.026	0.323	2.000
-400	4	0.007	0.253	-0.446	-0.101	-0.021	0.085	0.668
-400	6	0.011	0.326	-0.928	-0.082	0.029	0.234	0.465
Semi	auto	0.003	0.214	-0.560	-0.059	-0.001	0.093	0.420
		Growing nodules						
-600	3	0.017	0.712	-1.727	-0.195	0.016	0.294	2.000
-600	4	0.033	0.729	-1.483	-0.254	0.007	0.314	2.000
-600	6	0.136	0.757	-1.392	-0.129	0.054	0.271	2.000
-500	3	-0.044	0.709	-1.681	-0.174	-0.001	0.315	2.000
-500	4	-0.064	0.724	-1.717	-0.267	-0.021	0.377	2.000
-500	6	0.044	0.750	-1.714	-0.136	0.036	0.381	1.886
-400	3	-0.122	0.837	-2.000	-0.255	-0.026	0.323	2.000
-400	4	-0.138	0.883	-2.000	-0.190	-0.012	0.300	2.000
-400	6	0.027	0.758	-2.000	-0.100	0.075	0.351	2.000
Semi	auto	0.412	0.632	-0.265	-0.016	0.203	0.563	2.000

5 Discussion and Conclusion

It is difficult to assess the value of the three different methods that have been applied to the 50 nodule pairs of the VOLCANO'09 challenge, because we do not have access to information which nodule pairs are stable and which ones exhibit change. We therefore asked a radiologist to visually assess change in all pairs. Manually assessing small volume changes is hard for a human observer, and the radiologist repeatedly expressed his uncertainty, especially for nodules labeled as stable. It seems likely that some cases classified by the radiologist as stable are in reality slightly growing or shrinking. In some cases the automatic methods measured a change and the segmentation results of the methods were visually convincing.

Nevertheless, even in absence of any truth, we can conclude that the methods agree only moderately well with each other. The scatter plots in Figures 5 and 6 show large disagreement between the methods in a substantial amount of cases. Figure 1 shows that in our results, SEGM and SPHERE lead to a better separation between the stable and growing nodules than the REG method.

Table 3. Variability of ΔV outcome for the REG method using different values for threshold t and kernel size d. Stable and growth cases are treated separately. For each setting the average, standard deviation(SD), minimum, 1st quartile, median, 3rd quartile and the maximum value are reported. The row with label semi auto are the results from the semi-automatic experiments.

t	d	**Avg**	**SD**	**Min**	**q1**	**Median**	**q3**	**Max**
-400	3	0.033	0.208	-0.391	-0.116	0.037	0.148	0.415
-400	4	0.043	0.198	-0.203	-0.139	-0.022	0.137	0.446
-400	6	0.049	0.583	-0.542	-0.199	-0.050	0.181	2.000
-500	3	0.021	0.192	-0.365	-0.126	0.018	0.149	0.404
-500	4	0.030	0.208	-0.368	-0.138	0.082	0.142	0.433
-500	6	0.050	0.237	-0.332	-0.099	0.054	0.136	0.542
-600	3	0.109	0.337	-0.212	-0.112	0.054	0.188	1.308
-600	4	0.094	0.343	-0.379	-0.115	0.054	0.167	1.308
-600	6	0.104	0.326	-0.553	-0.058	0.093	0.189	1.276
Semi	auto	-0.033	0.209	-0.542	-0.165	-0.050	0.120	0.390
		Growing nodules						
-400	3	0.116	0.575	-0.851	-0.167	-0.040	0.138	2.000
-400	4	0.103	0.582	-0.918	-0.150	-0.073	0.140	2.000
-400	6	0.106	0.558	-0.897	-0.215	-0.023	0.233	1.641
-500	3	0.059	0.377	-0.517	-0.105	-0.024	0.110	1.470
-500	4	0.088	0.335	-0.722	-0.084	0.104	0.212	0.970
-500	6	0.126	0.313	-0.405	-0.044	0.102	0.242	0.936
-600	3	0.114	0.442	-0.856	-0.111	0.086	0.156	1.382
-600	4	0.143	0.392	-0.312	-0.092	0.080	0.203	1.329
-600	6	0.104	0.326	-0.553	-0.058	0.093	0.189	1.276
Semi	auto	0.206	0.423	-0.478	-0.067	0.103	0.377	1.641

Visual inspection of the results from the REG method indicated that the registration results were sometimes incorrect. Finding more stable settings for the registration method is therefore an important direction for future research, especially since this method was reported to work well in the work of Kabus et al. [4]. The final metric value reported by the registration software can be used to asses the quality of the registration and thus the quality of the measured change. This can be used to construct a fully automatic method, or to report the certainty of the system that the reported volume change is correct.

SEGM is very sensitive to the parameters used for the segmentation. These parameters vary from case to case and are can be hard to find, although in most cases it is a straightforward procedure to find good settings. SPHERE is not as sensitive to over-segmentation as SEGM, but it does rely heavily on the reported center point actually being in the center of the nodule. If the reported seed point of the nodule is not in the center, SPHERE will likely fail. An obvious solution to this problem is to include the determination of the sphere center point in the fitting procedure. Another weakness of SPHERE is the assumption of a spherical nodule, and spherical nodule growth. A more complicated model for nodule shape

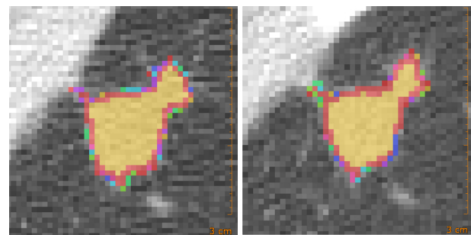


Fig. 2. Example of a successful segmentation by the SEGM method. The different colors in the image signify the amount of nodule in the voxel after sub-sampling, yellow means 100% of the voxel is nodule, red means 90%, pink means 80%, and purple means 70%. The other colors represent lower percentages of nodule in the voxel. The window center is -600 HU and the width of the window is 1600 HU (in all figures). The reported volumes by the SEGM method were 1776.9 mm^3 for scan 1 and 1442.5 mm^3 for scan 2, a ΔV of -0.19.

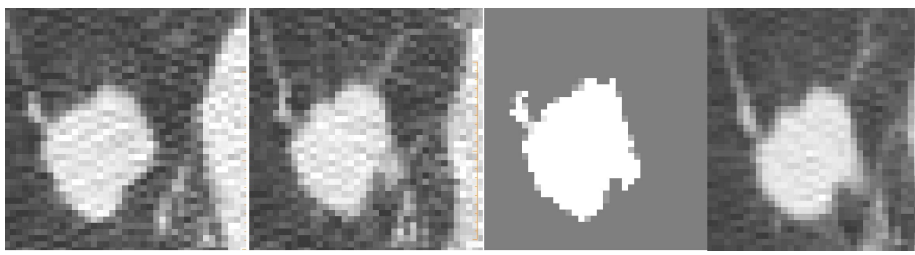
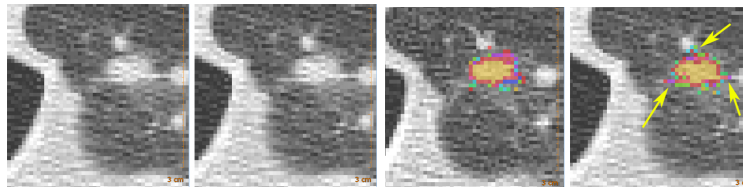


Fig. 3. Example of a successful segmentation by the REG method. From left to right; the nodule in the first scan, the transformed nodule after registration, the transformed segmentation and finally the nodule in the second scan. This case has been labeled shrinking by the human expert. The REG method reports a negative volume change ($V1 = 1511$ mm^3,$V2 = 1336$ mm^3 and ΔV = -0.116). The parameters for the segmentation used by REG are t = -400 HU and $d = 4$ mm).

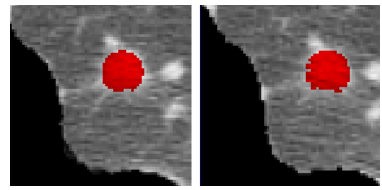
may improve the reliability of volume change assessment by SPHERE, but will also make the method much more complicated.

Finally, we note that it is not clear if measuring the relative volume change ΔV is the most important parameter to answer the question that is clinically the most relevant: is a nodule malignant or benign? The absolute volume change, or the nodule mass (change) may be better predictors, especially when used in conjunction with other features.

In conclusion, we have presented three simple methods for nodule volume change assessment and applied these to a public database provided by the VOLCANO'09 challenge. Although the methods produce visually convincing results in many cases, correlation between the methods is only moderately good, and the results do not show a clear separation between cases rated as stable versus growing by a radiologist.



(a) Nodule pairs (b) SEGM result

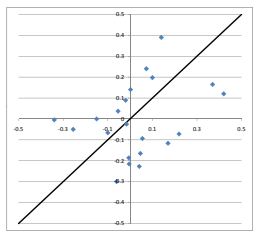
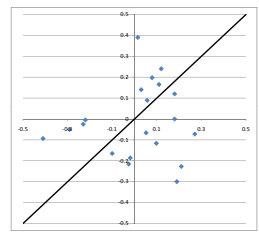
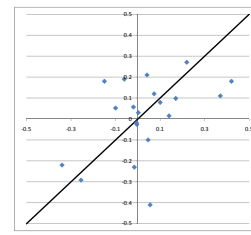


(c) SPHERE result

Fig. 4. An example where the SPHERE method ($t = -400\text{HU}, r = 0.5$) works better than the SEGM method ($t = -400\text{HU}, d = 3$ mm). The segmentation made by SEGM grows into some vessels in the second scan (indicated with arrows), but not in the other. This leads to an overestimation of nodule volume in scan 2. The SPHERE method clearly under-segments the nodule in both scans, but it does so in a consistent matter. The radiologist labeled this case as stable. SEGM reports growth ($V1 = 365.1$ mm^3, $V2 = 416.8$ mm^3, $\Delta V = 0.14$), SPHERE reports a nearly stable nodule ($\Delta V = 0.015$). Due to super-sampling figure 4(c) appears blurred.

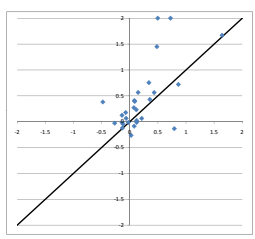
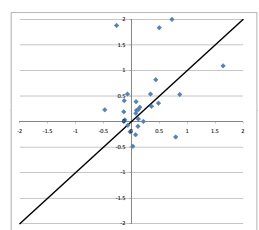
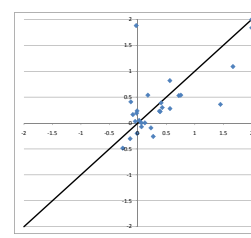
References

1. Wahidi, M.M., Govert, J.A., Goudar, R.K., Gould, M.K., McCrory, D.C., the American College of Chest Physicians: Evidence for the treatment of patients with pulmonary nodules: when is it lung cancer?: ACCP evidence-based clinical practice guidelines (2nd edition). Chest **132**(3 Suppl) (2007) 94S–107S
2. de Hoop, B., Gietema, H., van Ginneken, B., Zanen, P., Groenewegen, G., Prokop, M.: A comparison of six software packages for evaluation of solid lung nodules using semi-automated volumetry: what is the minimum increase in size to detect growth in repeated CT examinations. European Radiology **19**(4) (2009) 800–808
3. Reeves, A.P., Chan, A.B., Yankelevitz, D.F., Henschke, C.I., Kressler, B., Kostis, W.J.: On measuring the change in size of pulmonary nodules. IEEE Trans Med Imaging **25**(4) (2006) 435–450
4. Kabus, S., Müller, F., Wiemker, R., Fischer, B.: Robust lung nodule growth measurement by combining registration and segmentation. In: The First International Workshop on Pulmonary Image Analysis, Lulu.com 15–23
5. Sluimer, I.C., Prokop, M., Hartmann, I., van Ginneken, B.: Automated classification of hyperlucency, fibrosis, ground glass, solid and focal lesions in high resolution CT of the lung. Medical Physics **33**(7) (2006) 2610–2620
6. Kostis, W.J., Reeves, A.P., Yankelevitz, D.F., Henschke, C.I.: Three-dimensional segmentation and growth rate estimation of small pulmonary nodules in helical CT images. IEEE Trans Med Imaging **22**(10) (2003) 1259–1274
7. Klein, S., Staring, M.: Elastix. `http://www.isi.uu.nl/Elastix/` (Accessed March 30, 2009)

(a) REG versus SEGM (b) REG versus SPHERE (c) SEGM versus SPHERE

Fig. 5. Variability in outcome for stable nodules after semi-automatic experiments. Note that correlation is strong between segmentation and sphere, and particularly weak between registration and sphere.

(a) REG versus SEGM (b) REG versus SPHERE (c) SEGM versus SPHERE

Fig. 6. Variability in outcome for growing nodules after semi-automatic experiments. Note that correlation is strong between segmentation and sphere, and particularly weak between registration and sphere.