Automatic Segmentation of Pulmonary Segments From Volumetric Chest CT Scans

Eva M. van Rikxoort*, Bartjan de Hoop, Saskia van de Vorst, Mathias Prokop, and Bram van Ginneken, *Member, IEEE*

***Abstract*—Automated extraction of pulmonary anatomy provides a foundation for computerized analysis of computed tomography (CT) scans of the chest. A completely automatic method is presented to segment the lungs, lobes and pulmonary segments from volumetric CT chest scans. The method starts with lung segmentation based on region growing and standard image processing techniques. Next, the pulmonary fissures are extracted by a supervised filter. Subsequently the lung lobes are obtained by voxel classification where the position of voxels in the lung and relative to the fissures are used as features. Finally, each lobe is subdivided in its pulmonary segments by applying another voxel classification that employs features based on the detected fissures and the relative position of voxels in the lobe. The method was evaluated on 100 low-dose CT scans obtained from a lung cancer screening trial and compared to estimates of both interobserver and intraobserver agreement. The method was able to segment the pulmonary segments with high accuracy (77%), comparable to both interobserver and intraobserver accuracy (74% and 80%, respectively).**

***Index Terms*—Automatic, classification, computed tomography (CT), pulmonary lobes, pulmonary segments, segmentation.**

I. Introduction

THE advent of multidetector scanners that can acquire up to 64 slices per rotation has made high-resolution, submillimeter isotropic computed tomography (CT) imaging the modality of choice for analysis of the lungs. Modern CT scanners can generate a complete chest scan well within a single breath hold. A downside of this high-resolution imaging is that the analysis of these large amounts of data is very time consuming. This can be facilitated by computerized image analysis. Accurate segmentation is a prerequisite for automated analysis. One important task is the division of the lungs into anatomical regions.

The human lungs are divided into lobes [1]. The physical boundaries between the lobes are the pulmonary fissures. The lobes are further subdivided into segments that are defined based on bronchial supply. There are usually no physical boundaries between the segments. The pulmonary segments are a reference system for radiologists, pulmonologists, and surgeons to indicate the position of lesions in the lungs. Knowledge about the localization of lesions can guide bronchoscopy or surgery. In addition, the spatial distribution of a parenchymal disease can be of clinical importance. Pathologic abnormalities may be limited to one or several segments, in which case only the affected segments can be removed instead of complete lobes or lungs. Laros *et al.* [2] published a 30 year survey of 30 bronchiectatic patients for which more than 10 lung segments were removed. They concluded that as long as at least six normal lung segments were preserved, both the quantitative and qualitative functionality of the lungs did not deteriorate over the years. For patients with advanced emphysema, lung volume reduction surgery is a potentially valuable treatment [3]. For emphysema, the degree of deterioration can vary significantly between the left and right lung or between different lobes and segments.

Although segmentation of lobes and segments is useful in clinical practice, especially segments are underused because determining the segmental boundaries is time-consuming on 3-D CT data and usually hard on axial sections. In this paper we present a completely automatic method to identify both the pulmonary lobes and the pulmonary segments on 3-D CT data.

Several papers have been published about segmentation of pulmonary lobes. Zhang *et al.* [4] proposed an automatic lobe segmentation framework with atlas initialization for the lobar fissure segmentation. Next, a ridgeness measure was computed from 2-D slices to enhance fissure contrast, and fuzzy logic was used to extract the final lobar fissure positions. The extracted fissures were used to segment the lobes. Only the major fissure was extracted in both the left and right lung, so the right lung was divided into two lobes instead of three. It was shown that the similarity between the automatically found lobes and the manual lobe segmentation was high. The detection scheme was also applied to the minor fissure separating the upper and middle lobe in the right lung but this was not evaluated. Kuhnigk *et al.* [5] detected the lobar fissures using a distance transform to a vessel segmentation and the original CT values. Next, an interactive 3-D watershed transform was used to segment the lobes. They showed that the final result of their lobe segmentation did not vary substantially for different manual initializations. Wang *et al.* [6] proposed a method in which the major fissures were first segmented in a subset of sections, and subsequently a 3-D fissure interpolation was applied to obtain the fissure surface. Manual interaction was needed in 2.4% of sections. The segmentation results approximated the gold standard and were comparable to variation between human observers. The middle lobe in the right lung was not segmented since only the major fissure was extracted.

Manuscript received August 01, 2008; revised October 21, 2008. First published February 10, 2009; current version published March 27, 2009. This research was supported by the Dutch Organization for Scientific Research (NWO).. *Asterisk indicates corresponding author*

*E. M. van Rikxoort is with the Image Sciences Institute, 3584 CX Utrecht, The Netherlands (e-mail: eva@isi.uu.nl).

B. de Hoop, S. van de Vorst, and M. Prokop are with the Department of Radiology, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands.

B. van Ginneken is with the Image Sciences Institute, 3584 CX, Utrecht, The Netherlands.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2008.2008968

Automatic segmentation of the pulmonary segments has received little attention so far and the topic was listed in a recent survey [7] as a completely open research area. Kuhnigk *et al.* [5] proposed a method to extract pulmonary segments from 3-D CT data. After applying a lobe segmentation, they approximated the lung segments by assigning each lung voxel to the nearest point of the segmented bronchial tree in the same lobe. Since their bronchial tree segmentation includes an anatomic classification of the segmental branches this induces a partition of the lobes into segments. They evaluated the segment segmentation *in vitro* with two specimens of the left human lung. Based on this evaluation they concluded that for multidetector CT data more than 80% the voxels in the left lung were assigned to the correct segment for the two human left lung specimens. However, no evaluation was done on the real CT data nor on the right lung.

In this paper we will present the results of a completely automatic segment segmentation system and evaluate it on 100 low-dose CT scans obtained from a lung cancer screening trial. In this screening trial all abnormal findings are annotated by radiological experts and the segment they reside in is recorded. The performance of the automatic system will be compared to estimates of interobserver and intraobserver agreement.

In the next section, the pulmonary anatomy is briefly described, followed by a description of the data in Section IV. In Section III, the methods for lung segmentation, fissure segmentation, lobe segmentation, and segment segmentation are detailed. Section V gives the results of the automatic system and compares them to those of human observers. Section VI discusses the results and concludes.

II. Pulmonary Anatomy

The human lungs are divided into lobes and the physical boundaries between the lobes are the lobar fissures. The right lung consists of three lobes; the upper lobe, the middle lobe, and the lower lobe. The left lung consists of two lobes, and does not have a middle lobe. The equivalent of the middle lobe in the left lung is called the lingula. The lingula is occasionally considered a separate lobe in the left lung, however it is almost never separated from the upper lobe by a fissure [1]. In this paper, we will not consider the lingula to be a separate lobe in the left lung.

The lobes are separately supplied by the first subdivisions of the bronchial tree after the main bronchi (Fig. 1). Since also the vascular, nerve and lymphatic supply from the hilum to each lobe are mostly separated, the lobes function relatively independently within the lungs. The lobar fissures are often incomplete [8], in which case two lobes are (partly) connected. The middle row of Fig. 2 shows the lobes on cross-sectional slices from 3-D CT data.

The lobes are further subdivided into segments. There are almost never physical boundaries between the segments. Occasionally, an accessory fissure forms a physical boundary between two segments. The segments are defined based on supply from a tertiary branch of the bronchial tree. Based on the anatomy of the bronchial tree, 10 segments are defined in the right lung and eight in the left lung. Compared to the right lung, in the left lung no segment 7 is defined and segment 1 and 2 are considered a single segment, indicated as segment 1/2. In Fig. 1 a schematic drawing of the bronchial tree up to segmental level is shown. The bottom row of Fig. 2 shows the pulmonary segments on cross-sectional slices from 3-D CT data.

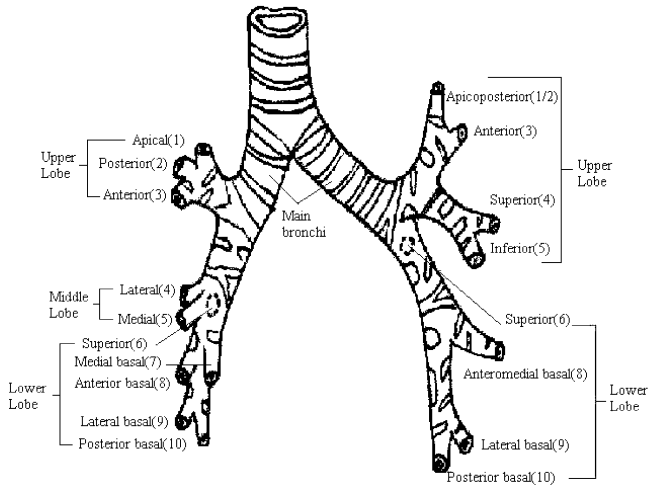

Fig. 1. Schematic drawing of the bronchial tree up to segmental level. Based on the anatomy of the bronchial tree 10 segments are defined in the right lung and eight in the left lung.

The segments form anatomical and functional regions of the lung parenchyma. However, the segments are less functionally independent than the lobes since veins tend to drain neighboring segments. Abnormalities are often localized in one or several segments.

III. Method

Before developing and evaluating the segment segmentation system, all scans were subsampled with a factor 2 in each direction using block averaging (the mean of eight voxels becomes the new voxel value) to reduce required computation times. All computations were performed on those subsampled data. The segment segmentation system consists of four steps. We start by segmenting the lungs. Next, the lobar fissures are segmented using a supervised approach. Based on the lung and fissure segmentations, the lobes are extracted. Finally, the segments are extracted per lobe. Since the fissure, lobe and segment segmentation are all based on classification, first a general section about the classifiers and features used is provided. Next, each of the four steps of the algorithm will be described.

A. Features and Classification

After the lungs are segmented, three similar approaches in which voxels are classified are used to perform fissure, lobe, and segment segmentation. This is an example of supervised learning in which we map input features to output for each voxel. This section describes the different features that are used and discusses the different classifiers used.

1) Gray-Scale Features: For the detection of the fissures inside the lungs two types of gray-scale features were employed; the eigenvalues of the Hessian matrix and gray-value information on different scales. Both sets of features are described in this section.

The Hessian matrix is a symmetric matrix that is composed from the six independent second-order derivatives. When performing an eigenvalue analysis of the Hessian matrix, the principal directions $(\hat{v}_0,\ \hat{v}_1,\ \hat{v}_2)$ in which the local second order structure can be decomposed are extracted. The corresponding eigenvalues $(|\lambda_0| \geq |\lambda_1| \geq |\lambda_2|)$ are real and denote the second-order derivative in the principal directions. One expects that at a fissure λ_0 is large, because when moving through the fissure the Hounsfield values change from low to high to low. The two other eigenvalues λ_1 and λ_2 are typically both low since the fissure itself is locally a plane with constant density. Thus the three eigenvalues together contain information that can be used to distinguish voxels on fissures from other voxels. To compute second order image derivatives, we used the standard procedure of convoluting the image with a Gaussian kernel with a specific scale σ, followed by obtaining derivatives with finite differences [9]. We used two scales, i.e., $\sigma = 1,3$ voxels.

The density itself is also an important characteristic, as voxels on fissures tend to have Hounsfield values in a specific range. We, therefore, added the gray value L at the same scales $\sigma = 1,3$ and also the voxel values themselves, indicated as L_0, which can be seen as observing the image at scale $\sigma = 0$ [9].

2) Position Features: For the lobe and segment segmentation, features based on positions inside and relative to objects will be employed. Two types of features were used, relative positions inside an object and positions relative to another object. Those features are described in this section.

To capture the position inside an object we computed so called cumulative positions inside the object. The cumulative positions indicate for each voxel in the object which percentage of object volume is above, next or behind it in each direction. Formally, the cumulative position $(\breve{x}, \breve{y}, \breve{z})$ of a point (x, y, z) inside an object is defined as follows:

$$\begin{aligned}
\breve{x} &= \frac{1}{V}\int_0^Z \int_0^Y \int_0^x I(x', y', z')dx'dy'dz' \\
\breve{y} &= \frac{1}{V}\int_0^Z \int_0^y \int_0^X I(x', y', z')dx'dy'dz' \\
\breve{z} &= \frac{1}{V}\int_0^z \int_0^Y \int_0^X I(x', y', z')dx'dy'dz' \qquad (1)
\end{aligned}$$

where V is the total number of voxels in the object, I is the segmentation image which is 1 inside the object and 0 elsewhere, and X, Y, and Z are the dimensions of the image. Use of the cumulative position maps locations inside the object to a standardized coordinate system, in the sense that voxels that are at a similar spatial location inside the object in different scans will get approximately the same coordinates.

Next to position inside an object, the position relative to another object can be employed in classification. To capture this relative position the direction to the closest point on the object was computed. This direction is normalized to unit length and the distance to the closest point on the object is added as a feature since that information is lost by normalizing the directional features. This results in four features: the direction in the x, y, and z direction and the distance. Clearly there are other ways in which this relative position can be encoded. We tried a number of different possibilities in pilot experiments on the training data and found our particular choice to work well in the context of the complete system.

3) Classification: The fissures, lobes, and segments are segmented using a supervised approach, i.e., a system is constructed using sample input and output data and classifiers from pattern recognition theory [10]. In supervised learning two stages can be distinguished: a training stage in which the system is developed, and a test stage in which the system is applied to previously unseen data. In the training stage, a number of voxels are sampled from training images, a set of features is calculated for each voxel and a classifier is trained [10]. In order to be able to train the classifier, a ground truth is required which gives the class label (i.e., belonging to the class to be segmented or not) for each voxel.

The choice of a classifier depends on the application. The performance of four classifiers was evaluated for each application in pilot experiments on the training data, i.e., the linear discriminant classifier (LDC), the quadratic discriminant classifier (QDC), the nearest mean classifier (NMC), and the *k*-nearest neighbor classifier (KNNC) [10]. The classifier that performed best in those pilot experiments was chosen as the classifier for the specific application.

For the task of fissure enhancement KNNC performed best. LDC performed best for both the lobe and segment segmentation task. LDC assumes a Gaussian distribution for the samples of each class and equal covariance matrices for each distribution [10]. The different means and the single covariance matrix that determine the discriminant were estimated from the training data using maximum likelihood.

B. Lung Segmentation

The lung fields were segmented with an automatic 3-D algorithm comparable to the algorithms proposed by Sluimer *et al.* [11] and Hu *et al.* [12]. This algorithm consists of the following steps.

1) First a structure comprising the trachea and the main parts of the bronchial tree is found using region growing. A seed point for this structure is determined by searching for a connected region on axial slices below -950 HU that matches certain size, location, and shape criteria. Starting at the top, each slice is examined until a suitable region is found. The seed is the point with the lowest Hounsfield value within this region. From this seed, the trachea and main stem bronchi are grown using explosion controlled region growing. This means that the threshold applied is slightly increased in each iteration and the grown structure is frozen when its size increases by a factor two compared to the previous threshold. Such a volume increase (the explosion) indicates that the bronchi have merged with the lung parenchyma. The first threshold used is -950 HU. From the determined structure, the point with the lowest Hounsfield value is taken as seed point for the next step.
2) From the new seed point region growing is applied to find the lungs. Optimal thresholding as described by Hu *et al.* [12] is used to determine the upper threshold for this region growing operation.

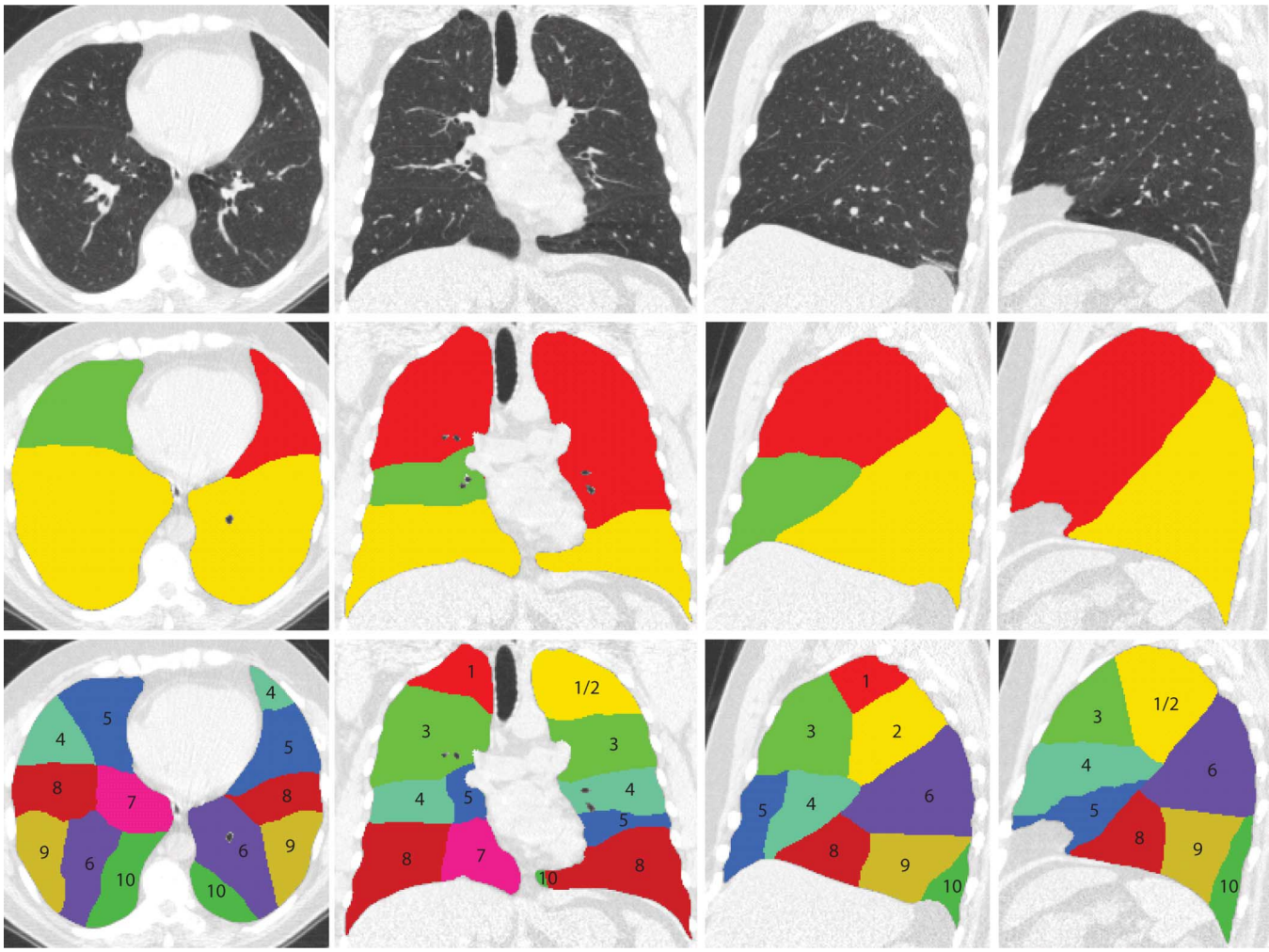

Fig. 2. Illustration of pulmonary lobar and segmental anatomy from actual results of the different steps of the segment segmentation algorithm on a randomly selected scan. From left to right: an axial slice, a coronal slice, a sagittal slice of the right lung and a sagittal slice of the left lung. The top row shows slices of the original scan. The second row shows the result of the lobe segmentation, in the bottom row the corresponding segment segmentation is shown.

3) After the lungs are grown, the trachea and bronchi found in the first step are removed from the results to obtain only the lungs. Sometimes only one connected component is found. In those cases, the left and right lung are only separated by a low contrast junction line that is the pleura. To separate the lungs in this case, dynamic programming in axial slices is applied, similar to Hu *et al.* [12].
4) As a final step for the lung segmentation, each lung is smoothed separately using 3-D hole filling and morphological closing with a spherical structuring element of size 11 to include vessels in the segmentation and smooth the borders.

Example results of the lung segmentation can be seen in Figs. 2 and 4.

C. Fissure Segmentation

After the lungs are segmented, the fissures in the lung are detected. Fissure segmentation is initiated by applying a supervised enhancement filter as described in [13]. The results of the enhancement filter will be for each voxel inside the lungs a probability that it belongs to a fissure. Next, the major and minor fissures are extracted from the enhanced image by grouping neighboring voxels into plates. In this section first the enhancement filter will be described followed by a description of the grouping process that leads to the final segmentation of the major and minor fissures.

1) Fissure Enhancement: The fissures inside the lungs are enhanced by applying a supervised enhancement filter [13]. To be able to train the fissure enhancement filter, examples of positive and negative voxels are needed. All voxels of the manually segmented fissures (see Section IV) were used as positive examples. An equal number of points elsewhere in the lungs were selected as negative examples. Since the voxels around the fissures are the most challenging negative voxels, half of the negative voxels were forced to be within five voxels of the manually drawn fissures, the other half was more than five voxels away.

For each voxel in the training data set, a set of features is calculated that describe certain characteristics of that voxel. The set of features is chosen in such a way that the characteristics that are important for the structure to be enhanced are represented in the features. Since the fissures are bright plates in the lungs the filter used the eigenvalues of the Hessian matrix

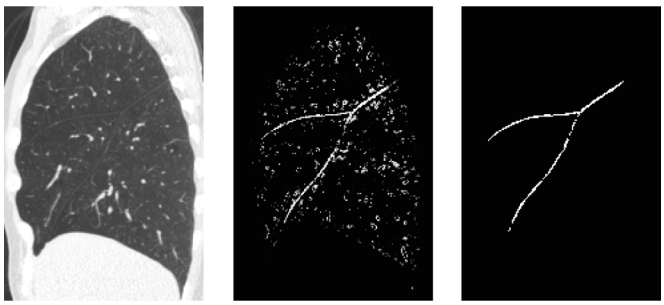

Fig. 3. Example of the fissure segmentation, (a) the original CT slice, (b) result of supervised enhancement, (c) the final fissure segmentation.

($|\lambda_0| \geq |\lambda_1| \geq |\lambda_2|$) and grey value information (L and L_0) as features (see Section III-A). All features except L_0 were calculated on two scales, i.e., $\sigma = 1, 3$ voxels. This results in nine features in total, $\lambda_0, \lambda_1, \lambda_2$ and L at two scales plus L_0.

Based on the features and class labels for each voxel in the training set, a KNNC was trained to be able to assign to unseen voxels a probability that they belong to a fissure. On the training data, the value of k was varied between 1 and 45 and found to be optimal at 15.

To apply the fissure enhancement filter as developed in the training stage described above to a test image, each voxel inside the lungs is classified by the trained classifier. The result of the fissure enhancement is for each voxel inside the lungs a probability that it belongs to a fissure. An example result of the enhancement filter can be seen in Fig. 3(b).

2) Fissure Segmentation: Since we are only interested in the major and minor fissures, i.e., the boundaries between the lobes, the enhanced fissures are converted to a segmentation of the major and minor fissures. To accomplish this, the probabilities resulting from the fissure enhancement (i.e., posterior probabilities) are first thresholded at 0.5 to eliminate all voxels with a low probability of being on a fissure. Next, since fissures are plates in the lungs, the remaining voxels are grouped into plates using a method inspired on the technique described in [14] and [15]. This method groups neighboring voxels based on their likelihood of constituting a plane given second-order image information.

As described in Section III-A the direction of the largest curvature is perpendicular to the plate and is described by the eigenvector $\hat{v}_0$ of the Hessian matrix. We employ $\hat{v}_0$ calculated at scale $\sigma = 1$ to group voxels belonging to the same plate. For a voxel v_a all voxels v_b within a distance d are considered to be on the same plate if 1) the directions of $\hat{v}_0$ are similar for both voxels and 2) the voxels belong to the same plate (and not to parallel plates). The first condition is checked by taking the inner product of the normalized eigenvectors $\hat{v}_0$ at the locations of v_a and v_b; if v_a and v_b have similar orientation the product will be close to one. The second condition is checked by taking the outer product of the normalized difference vector $\hat{w}$ between the locations of v_a and v_b and $\hat{v}_0$ of v_a; if the voxels are on the same plate the outer product will be close to one.

To discard accessory fissures and spurious detections only plates with a sufficient size were retained. The algorithm as described above contains four parameters that need to be set: 1) the maximum distance d between the two voxels under consideration, 2) a threshold T_v on the product of the two eigenvectors, the closer T_v is to 1 the less curvature is allowed in the resulting plates, 3) a threshold T_w on the product between the difference vector and the eigenvector of the grouped vector, the closer T_w is to 1 the less chance that two parallel plates are considered the same plate, and 4) the minimum number of voxels in a plate to be considered a major or minor fissure. The final result of the fissure segmentation is a set of plates in the lungs which form the major and minor fissures. An example of the final segmentation result can be seen in Fig. 3(c).

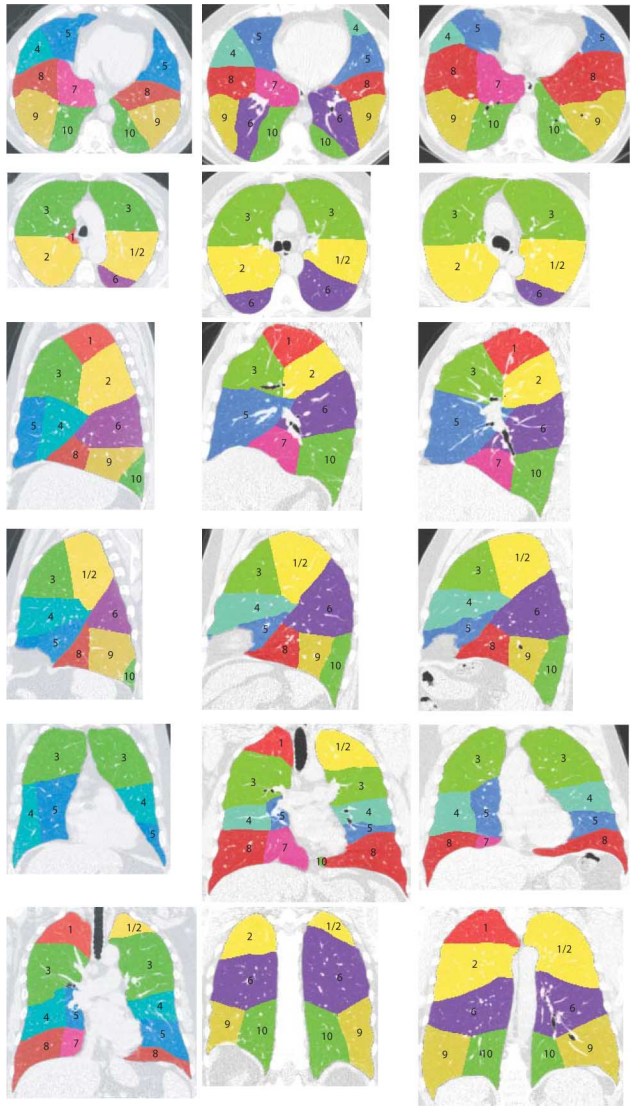

Fig. 4. Typical results of the automatic segment segmentation for 3 (the columns) randomly selected scans on different cross-sectional slices. Slices have been chosen in such a way that all segments are shown in each direction at least once.

D. Lobe Segmentation

The lobes can not be directly obtained from the fissure segmentations; the fissures are often incomplete or invisible on the CT scan or may not be detected correctly by our algorithm. To be able to cope with this, we segment the lobes using a supervised approach. Lobe segmentation is applied to each lung separately.

To train the lobe segmentation, voxels in the training data were used for which a segment label is known which is converted to a lobe label. See Section IV for a detailed description of the labeled voxels.

Next, for each voxel a set of features is calculated which are important for lobe segmentation. Based on the data available, i.e., the lung and fissure segmentations, the most important feature for a voxel is its position relative to the fissures. We captured this relative position to the fissures using the direction and the distance to the closest point on the fissures (see Section III-A). This results in four features; the direction in the x, y, and z direction and the distance.

In cases where the fissure segmentation is not complete, additional information can be obtained from the position inside the lungs. To capture the position inside the lung we employed the cumulative position in the lung as features (see Section III-A). The advantage of using cumulative positions is that locations inside the lungs are mapped to a standardized coordinate system. With $\breve{x}$,$\breve{y}$, and $\breve{z}$ added to the feature set, seven features were used in total.

To be able to apply the lobe segmentation to test data, the feature set created using the training data is used to train a LDC. The result of the lobe segmentation will be for each voxel a set of probabilities that it belongs to each lobe. To obtain the final segmentation result, the posterior probabilities were first postprocessed by a Gaussian blurring with a scale $\sigma = 2$ voxels. This has the effect of pooling local evidence and smoothing the lobe borders. Next, each voxel is assigned to the lobe with the highest probability. A result of the lobe segmentation can be seen in the second row of Fig. 2.

E. Segment Segmentation

As a final step, each voxel is assigned to a pulmonary segment. Since there are typically no physical boundaries between the segments we must use more indirect evidence and we attempted to exploit the fact that the spatial arrangement between the segments in a lobe is relatively constant. Segment segmentation is performed per lobe using a voxel classification approach similar to the one described in Section III-D. As training data for the segment segmentation the locations from the training data for which a segment label is known were used (see Section IV for the description of the training data).

Because for the assignment of a segment label to a voxel its position inside the lobe is most important, the cumulative position features are used as features with the object I in (1) being the lobe segmentation. Next to the position inside the lobe, the position relative to the lobar fissures adds useful additional information. Therefore, the distance and direction to the closest point on the fissures were added as features, which makes seven features in total.

A LDC was trained to assign to each voxel in a test image a probability that it belongs to each segment. The posterior probabilities resulting from the classification were blurred with a scale $\sigma = 1$ voxel before each voxel is assigned to the segment with the highest posterior probability. Examples of the segment segmentation can be seen in the bottom row of Fig. 2 and in Fig. 4.

IV. Data

For this study data was taken from the NELSON study, a Dutch lung cancer screening trial [16] with low dose CT (30 mAs at 120 kV for patients weighing ≤ 80 kg and 30 mAs at 140 kV for those weighing over 80 kg). Data was acquired on Mx8000IDT or Brilliance-16 CT scanners (Philips Medical Systems, Cleveland, OH) in about 12 s in spiral mode with 16×0.75 mm collimation and 15-mm table feed per rotation $(\text{pitch} = 1.3)$. Axial images of 1.0 mm thickness at 0.7 mm increment were reconstructed using a moderately soft kernel (Philips "B") with the smallest field-of-view that included the outer rib margins at the widest dimension of the thorax. All scans were reconstructed to 512×512 matrices.

In the lung cancer screening trial, all abnormal findings (mostly pulmonary nodules) are annotated by radiologists and for each finding the pulmonary segment in which it resides is recorded. From all scans obtained in our hospital between January 2004 and May 2006 containing more than four findings, 600 scans from different subjects were randomly selected. From these, 500 scans with 3439 findings were used as a training set for the segmentation system. The remaining 100 scans with 697 findings were set aside and used for testing. They were not used in any way during system development. For this set, a second radiologist assigned a pulmonary segment to all findings. Moreover the radiologist that annotated the findings for the screening data repeated the assignment, this second assignment was done in a blind fashion, i.e., without having access to the results of the first reading. This allows us to compare the performance of the automatic system to estimates of both the interobserver and intraobserver agreement.

From the three annotations that exist for the evaluation data, a consensus set was constructed which contained only those lesions for which all three annotations agree on the segment label. In total, the observers agreed on 467 out of 697 findings (67%), 275 of those lesions were in the right lung and 192 in the left lung. So in total there are four sets of annotations available for evaluation of the automatic system

1) The annotations from the screening trial, which will be referred to as the reference.
2) The annotations of the second radiologist, from which an estimate of the interobserver variability is computed.
3) The second reading of the reference radiologist, from which an estimate of the intraobserver variability is computed.
4) The consensus set which contains only those lesions on which all three annotations agree.

The accuracy compared to the reference will be computed for the automatic system as well as for the human observers. In addition the accuracy for the automatic system on the consensus set will be calculated. The number of findings in each segment for each lung are given in Table I for the 500 training scans, the reference annotations of the 100 test scans and the consensus data.

For the fissure segmentation that is applied as one of the parts of the segment segmentation (see Section III) a set of 13 normal dose (120 kV, 100–150 mAs) inspiration CT chest scans of 13 different patients were randomly selected from clinical practice at our hospital as training data. This data was acquired

TABLE I
NUMBER OF FINDINGS RECORDED IN EACH SEGMENT PER LUNG IN THE 500 SCANS USED FOR TRAINING THE SYSTEM, IN THE 100 SCANS USED FOR EVALUATING THE SYSTEM (BASED ON THE REFERENCE ANNOTATIONS), AND IN THE CONSENSUS DATA

Right Lung		Left Lung	
500 training scans			
segment	**# of lesions**	**segment**	**# of lesions**
1	154		
2	239	1/2	219
3	221	3	173
4	215	4	128
5	278	5	100
6	305	6	242
7	19		
8	168	8	200
9	214	9	311
10	126	10	127
100 evaluation scans			
segment	**# of lesions**	**segment**	**# of lesions**
1	30		
2	39	1/2	44
3	61	3	38
4	49	4	30
5	53	5	22
6	67	6	61
7	1		
8	29	8	27
9	41	9	44
10	33	10	28
consensus data			
segment	**# of lesions**	**segment**	**# of lesions**
1	27		
2	26	1/2	31
3	52	3	21
4	34	4	11
5	15	5	8
6	56	6	52
7	0		
8	16	8	19
9	26	9	24
10	23	10	26

on the same scanners and reconstructed with the same kernel as the data from the screening trial described above. In these scans a human observer manually indicated the lobar fissures in every fourth coronal slice. Segmenting fissures was performed by clicking points on the fissure; between two points, a straight line was automatically drawn.

V. EXPERIMENTS AND RESULTS

The lobe and segment segmentation systems were trained using the 500 scans in the training data set for which one annotation exists for each lesion. Next, the complete system was applied to the 100 scans in the evaluation set for which three annotations exist; the reference, the interobserver and the intraobserver annotations. To obtain the segmentation of the images in their original size, the segmentation results were super-sampled to the original resolution using nearest neighbor interpolation. As described in Section III-C2 four parameters need to be set for the task of fissure segmentation. For the current application d was set to 3, T_v and T_w were both set to 0.985 and the minimum number of voxels was set to 500. These settings were determined in pilot experiments. In Fig. 4 typical results of the segment segmentation are shown for three scans in different cross-sectional slices. The slices were chosen in such a way that each segment is visible at least once in each direction.

In Table II the confusion matrices for the left and right lung as produced by the automatic system and the estimates of interobserver and intraobserver variability are given as compared to the reference for the complete data set of 100 scans (697 findings). In addition, computer performance on the consensus data only (467 findings) is listed. For the complete test data set, the accuracy of the automatic system was 81% for the left lung and 74% for the right lung. The total accuracy for the automatic segment segmentation was 77%. The interobserver agreement was 75% for the left lung and 74% for the right lung, the intraobserver agreement was 81% on the left lung and 79% for the right lung. In the last section of Table II the confusion matrices for the automatic system as compared to the consensus data are shown. The accuracy for both lungs in this case was 88%. The lung segmentation that was applied as a first step was very successful; all 697 findings were assigned to the correct lung.

The lobe segmentation that is applied as part of the segment segmentation system was evaluated using the same data as for the evaluation of the segment segmentation. In Table III the confusion matrices for the lobe segmentations for each lung are given. For the left lung, the total accuracy of the lobe segmentation as compared to the reference was 97%, for the right lung this was 90%. The interobserver agreement for the lobe segmentation was 97% for the left lung and 87% for the right lung, the intraobserver agreements were 96% and 88%, respectively. On the consensus data the accuracy for the lobe segmentation of the automatic system was 98% for the left lung and 95% for the right lung.

VI. DISCUSSION AND CONCLUSION

We have presented the first fully automatic approach to extraction of pulmonary segments from chest CT data. The results indicate that the system is able to assign a segment label to each voxel inside the lungs with an accuracy comparable to that of human observers.

Not only are the overall results comparable to those of human observers, the confusion matrices as presented in Section V show that most of the errors produced by the automatic system are similar to disagreement between expert human observers. The system performs slightly better than an estimate of the interobserver agreement and slightly worse than the estimated intraobserver agreement. One should note that the human observers assigned segment labels to the lesions in the 100 test scans in one session. Therefore, it is possible that they were more consistent in their labelling than in a normal clinical setting.

In general, the automatic system performed better for the left lung than for the right lung. Inspection of the confusion matrices shows that this is mainly due to relatively low accuracies obtained for segments 8, 9, and 10 in the right lung (56%, 66%, and 71%, respectively). This is, however, comparable to performance of the human observers, especially the interobserver agreement. The second observer labelled findings in segments 8, 9, and 10 with an accuracy of 59%, 67%, and 70%, respectively. The human observers and the automatic system assigned the lesions that were supposed to be in those segments to the same segments; segment 8 was mainly confused with segment 4, segment 9 was mainly confused with segment 6 and 8, and segment

TABLE II

CONFUSION MATRICES FOR THE SEGMENT SEGMENTATION IN EACH LUNG. NUMBERS ARE PERCENTAGES. THE COLUMNS INDICATE THE CLASSIFICATION, THE ROWS INDICATE THE GROUND TRUTH. SO, FOR EXAMPLE ENTRANCE $(0, 1) = 2$ INDICATES THAT OF ALL FINDINGS THAT SHOULD BE, ACCORDING TO THE REFERENCE ANNOTATIONS, IN SEGMENT 1, 2% ARE CLASSIFIED AS BEING IN SEGMENT 2

Right Lung

Automatic System on full test data (697 findings)

Reference \ *Classified segment*	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	**90**	10	0	0	0	0	0	0	0	0
2	23	**63**	11	0	0	3	0	0	0	0
3	2	0	**79**	5	14	0	0	0	0	0
4	0	0	4	**79**	13	4	0	0	0	0
5	0	0	7	7	**80**	4	0	0	2	0
6	0	6	0	5	5	**77**	0	2	0	5
7	0	0	0	0	0	0	**100**	0	0	0
8	0	0	0	19	7	0	7	**56**	11	0
9	0	0	0	0	0	12	0	22	**66**	0
10	0	0	0	0	0	13	3	0	13	**71**

Inter Observer on full test data (697 findings)

Reference \ *Classified segment*	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	**93**	7	0	0	0	0	0	0	0	0
2	15	**67**	13	0	0	5	0	0	0	0
3	2	1	**95**	0	2	0	0	0	0	0
4	0	2	0	**82**	4	12	0	0	0	0
5	0	0	34	24	**30**	8	2	2	0	0
6	0	4	0	0	0	**93**	0	0	0	3
7	0	0	0	0	0	**100**	0	0	0	0
8	0	0	3	10	3	14	4	**59**	7	0
9	0	0	0	0	0	19	0	12	**67**	2
10	0	0	0	0	0	27	0	0	3	**70**

Intra Observer on full test data (697 findings)

Reference \ *Classified segment*	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	**90**	7	3	0	0	0	0	0	0	0
2	8	**87**	0	0	0	5	0	0	0	0
3	0	7	**88**	2	3	0	0	0	0	0
4	0	2	0	**72**	14	8	0	2	0	2
5	0	0	26	11	**55**	4	0	2	2	0
6	0	5	1	5	0	**85**	0	1	1	2
7	0	0	0	0	0	0	**100**	0	0	0
8	0	0	0	10	4	0	0	**79**	7	0
9	0	0	0	0	0	10	0	12	**71**	7
10	0	0	0	0	0	9	0	0	3	**88**

Automatic system on consensus data only (467 findings)

Reference \ *Classified segment*	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	**93**	7	0	0	0	0	0	0	0	0
2	16	**84**	0	0	0	0	0	0	0	0
3	2	0	**82**	4	12	0	0	0	0	0
4	0	0	0	**91**	3	3	0	0	3	0
5	0	0	0	7	**93**	0	0	0	0	0
6	0	1	0	2	4	**89**	0	0	0	4
7	0	0	0	0	0	0	**0**	0	0	0
8	0	0	0	0	0	0	6	**81**	13	0
9	0	0	0	0	0	0	0	15	**85**	0
10	0	0	0	0	0	0	5	0	5	**90**

Left Lung

Automatic System on full test data (697 findings)

Reference \ *Classified segment*	**1/2**	**3**	**4**	**5**	**6**	**8**	**9**	**10**
1/2	**88**	10	2	0	0	0	0	0
3	5	**81**	14	0	0	0	0	0
4	0	7	**59**	24	10	0	0	0
5	4	0	10	**86**	0	0	0	0
6	0	0	8	0	**81**	11	0	0
7								
8	0	0	0	0	5	**77**	18	0
9	0	0	0	0	2	12	**84**	2
10	0	0	0	0	3	0	4	**93**

Inter Observer on full test data (697 findings)

Reference \ *Classified segment*	**1/2**	**3**	**4**	**5**	**6**	**8**	**9**	**10**
1/2	**91**	9	0	0	0	0	0	0
3	24	**68**	5	0	3	0	0	0
4	3	27	**47**	10	10	0	3	0
5	5	4	32	**50**	9	0	0	0
6	0	0	0	0	**92**	0	3	5
7								
8	0	0	0	3	7	**79**	11	0
9	0	0	0	0	7	32	**57**	4
10	0	0	0	0	7	0	0	**93**

Intra Observer on full test data (697 findings)

Reference \ *Classified segment*	**1/2**	**3**	**4**	**5**	**6**	**8**	**9**	**10**
1/2	**100**	0	0	0	0	0	0	0
3	21	**68**	5	3	3	0	0	0
4	0	20	**57**	10	10	0	3	0
5	5	0	36	**50**	9	0	0	0
6	2	0	2	0	**92**	2	2	0
7								
8	0	0	4	4	0	**74**	18	0
9	0	0	0	0	0	18	**77**	5
10	0	0	0	0	0	0	0	**100**

Automatic system on consensus data only (467 findings)

Reference \ *Classified segment*	**1/2**	**3**	**4**	**5**	**6**	**8**	**9**	**10**
1/2	**90**	10	0	0	0	0	0	0
3	0	**90**	10	0	0	0	0	0
4	0	0	**64**	36	0	0	0	0
5	0	0	0	**100**	0	0	0	0
6	0	0	5	0	**85**	10	0	0
7								
8	0	0	0	0	6	**82**	12	0
9	0	0	0	0	4	0	**96**	0
10	0	0	0	0	4	0	4	**92**

10 was mainly confused with segment 6. Note that all those pairs of segments are neighboring. In addition, in the right lung segment 2 was often confused for segment 1 and 3 (15% and 11%, respectively) by the automatic system. This is comparable to the interobserver agreement (15% and 13%, respectively).

For both lungs, distinguishing segment 4 and 5 was problematic for both the human observers and the automatic system. For segment 4 the automatic system performed comparably to the human observers in both lungs, for segment 5 however, the automatic system had a much higher accuracy, compared to our reference, than the human observers. The human observers not only confused segments 4 and 5 with each other, segment 5 in the right lung and segment 4 in the left lung were often confused with segment 3 in the specific lungs. In the right lung this was probably due to incompleteness or absence of the minor fissure separating the upper and middle lobe. Since there is usually no minor fissure in the left lung it is even more difficult to assign the correct segmental labels to positions near segmental borders. The fact that different segments were confused with segment 3 in each lung is due to the different orientation of segment 4 and 5 in both lungs, as can be seen from Fig. 4.

When only the consensus data was used for evaluation, computer results improved to an accuracy of 88% for both lungs. For the consensus data the main problems were the same as for the total evaluation; in the right lung segments 8 and 9 were the

TABLE III
CONFUSION MATRICES FOR THE LOBE SEGMENTATION IN EACH LUNG, THE NUMBERS ARE PERCENTAGES. THE COLUMNS INDICATE THE CLASSIFICATION, THE ROWS INDICATE THE GROUND TRUTH. **1** INDICATES THE UPPER LOBE, **2** INDICATES THE MIDDLE LOBE AND **3** THE LOWER LOBE

		Right Lung				Left Lung		
Automatic System								
		Classified lobe				*Classified lobe*		
		1	**2**	**3**		**1**	**2**	**3**
Reference	**1**	**91**	9	0	**1**	**98**		2
	2	5	**90**	5	**2**			
	3	2	8	**90**	**3**	3		**97**
Inter Observer								
		Classified lobe				*Classified lobe*		
		1	**2**	**3**		**1**	**2**	**3**
Reference	**1**	**90**	1	9	**1**	**95**		5
	2	18	**70**	12	**2**			
	3	2	3	**95**	**3**	1		**99**
Intra Observer								
		Classified lobe				*Classified lobe*		
		1	**2**	**3**		**1**	**2**	**3**
Reference	**1**	**89**	2	9	**1**	**95**		5
	2	14	**76**	10	**2**			
	3	2	4	**94**	**3**	3		**97**
Automatic system on consensus data								
		Classified lobe				*Classified lobe*		
		1	**2**	**3**		**1**	**2**	**3**
Reference	**1**	**92**	8	0	**1**	**100**		0
	2	0	**96**	4	**2**			
	3	1	3	**97**	**3**	3		**97**

most difficult and in the left lung segment 4 was most often confused with segment 5.

Table I shows that the distribution of findings over the different segments is not uniform. In particular, there is only one finding in segment 7 of the right lung in the test data, and there is no consensus for this finding among the observers. Segment 7 is relatively small and pulmonary nodules are apparently relatively rare in that part of the lungs.

To obtain a qualitative evaluation of the automatic segment segmentation an expert radiologist visually inspected the pulmonary segments extracted by the automatic system for several scans and found the result to be convincing. The quantitative evaluation performed in this paper is based on random points in a large set of scans and reflects the way the segments are used by radiologists; as a reference for precise localization of lesions in the lungs. The results from Tables II and III in combination with the images in Figs. 2 and 4 show that the segmentation of the pulmonary segments in the evaluation set was successful.

The performance of the lobe segmentation was evaluated using the same data and was comparable to both the intraobserver and interobserver agreement. For the middle lobe in the right lung, the automatic system performed better than the human observers. This is probably due to the common absence or incompleteness of the minor fissure separating the upper and middle lobe. This is illustrated by the performance of the system on the consensus data; the accuracy of the automatic system for the middle lobe improved to 96%, which means that only two lesions were assigned to the wrong lobe. Both those lesions were assigned to the upper lobe instead of the middle lobe.

We inspected all lesions that were assigned to the wrong lobe for the consensus data. For the left lung in total three out of 193 lesions were misclassified. We found that all those points were

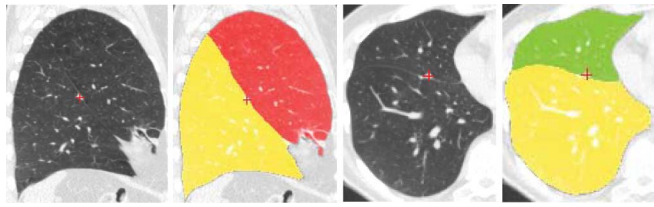

Fig. 5. Example of PFOs and the corresponding lobe segmentation as produced by the automatic lobe segmentation. The red cross indicates the point marked in the nodule.

perifissural opacities (PFOs) [17], [18], i.e., nodules attached to a fissure. It is difficult to assign the correct label to such findings. A point that is marked inside the nodule might be on the other side of the fissure than the main body of the nodule. In Fig. 5 two examples of PFOs and the corresponding lobe segmentations are given. For the right lung 14 out of 275 lesions in total were misclassified of which six were PFOs. Of the other errors, three positions were on the border between two lobes where no fissure was found, four errors were due to a segmentation error in the lobe segmentation and one error was a nodule on the border of the lungs which was not included in our lung segmentation.

All scans were subsampled with a factor 2 in each direction to reduce the amount of computation time required. The algorithm was implemented in C++, and the code was not optimized. It takes on average around 10 min to process a complete scan on a 3-GHz Pentium 4 machine. This time is subdivided into 1 min for the lung segmentation, 4 min for the fissure segmentation, 2 min for the lobe segmentation, and finally 3 min for the segment segmentation.

It is important to note that in the test data used in the current study no significant pathologic abnormalities were present except for emphysema, since the data was taken from a lung cancer screening trial. The presence of emphysema does not appear to have a detrimental effect on system performance. We expect the system to be fairly robust against pathologic abnormalities due to the combination of positional and fissure features. However, when an abnormality locally resembles a fissure or if the lung segmentation fails due to the presence of large abnormalities, the complete system is likely to produce poor results. A successful fissure enhancement filter is essential for calculating the distance to the fissures. Visual inspection showed that fissures that were visible on the CT data were always accurately detected. The fact that the training data for the fissure filter consisted of clinical dose scans while the test data here are low dose screening scans did not appear to have a detrimental effect on the fissure filter. Fig. 6 shows an example result of the automatic segmentation on a scan containing substantial pathologic abnormalities. A quantitative evaluation of automatic segmentation of pulmonary segments in scans that contain abnormalities is an important topic for future work.

Since fissures are often more than one voxel in width, some of the voxels that are used as negative examples will actually be fissure voxels. Due to the small amount of voxels for which this is the case compared to the total number of negative voxels this will not have much influence on the performance of the system.

Although the automatic segmentation of the pulmonary segments as presented in this paper is largely successful, the system

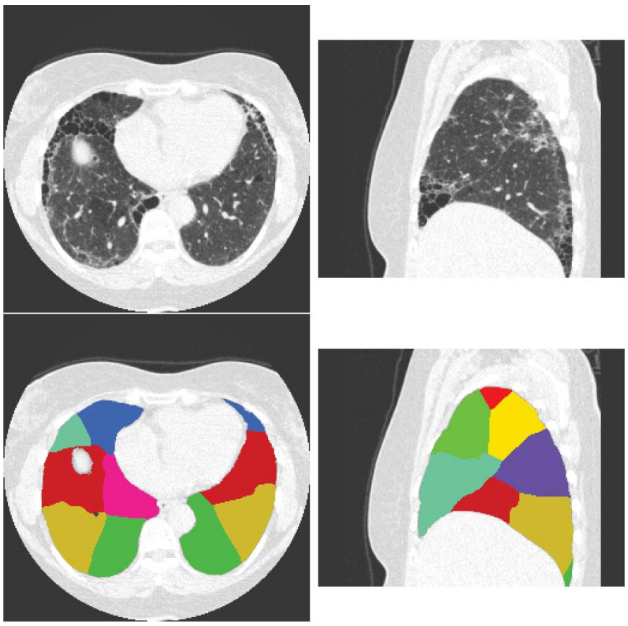

Fig. 6. Example result of the segment segmentation on a scan containing pathologic abnormalities. Despite the presence of abnormalities the result of the segment segmentation is comparable to the results on the screening data.

might be further improved by including a segmentation of the bronchial tree and the arterial system. A scheme in which the bronchial tree and arterial system provide additional evidence in cases were the current system fails might be more robust. It is, however, not clear how the substantial variations that exist in bronchial tree anatomy between subjects influence the segmental boundaries.

In conclusion, this study has shown that it is feasible to assign with high accuracy pulmonary segment labels to voxels on the basis of a small set of features that encode a voxel's position in the lungs, the lobes and relative to automatically detected fissures. The system was evaluated using random positions in a large set of scans from a screening trial. The results showed that the automatic system performed well, with an accuracy comparable to that of human experts.

REFERENCES

[1] M. Prokop and M. Galanski, , M. Prokop and M. Galanski, Eds., *Spiral and Multislice Computed Tomography of the Body*. Stuttgart, Germany: Thieme, 2003.

[2] C. D. Laros, J. M. M. van den Bosch, C. J. J. Westermann, P. G. M. Bergstein, R. G. J. Vanderschueren, and P. Knaepen, "Resection of more than 10 lung segments; a 30-year survey of 30 bronchiectatic patients," *J. Thoracic Cardiovascular Surg.*, vol. 95, no. 1, pp. 119–123, 1988.

[3] N. E. T. T. R. Group, "Patients at high risk of death after lung-volume-reduction surgery," *N. Eng. J. Med.*, vol. 345, no. 15, pp. 1075–1083, 2001.

[4] L. Zhang, E. A. Hoffman, and J. M. Reinhardt, "Atlas-driven lung lobe segmentation in volumetric x-ray CT images," *IEEE Trans. Med. Imag.*, vol. 25, no. 1, pp. 1–16, Jan. 2006.

[5] J.-M. Kuhnigk, V. Dicken, S. Zidowitz, L. Bornemann, B. Kuemmerlen, S. Krass, H.-O. Peitgen, S. Yuval, H.-H. Jend, W. S. Rau, and T. Achenbach, "New tools for computer assistance in thoracic CT part 1. Functional analysis of lungs, lung lobes and bronchopulmonary segments," *Radiographics*, vol. 25, pp. 525–536, 2005.

[6] J. Wang, M. Betke, and J. P. Ko, "Pulmonary fissure segmentation on CT," *Med. Image Anal.*, vol. 10, pp. 530–547, 2006.

[7] I. C. Sluimer, A. M. R. Schilham, M. Prokop, and B. van Ginneken, "Computer analysis of computed tomography scans of the lung: A survey," *IEEE Trans. Med. Imag.*, vol. 25, no. 4, pp. 385–405, Apr. 2006.

[8] K. Hayashi, A. Aziz, K. Ashizawa, H. Hayashi, K. Nagaoki, and H. Otsuji, "Radiographic and CT appearances of the major fissures," *Radiographics*, vol. 21, no. 4, pp. 861–874, 2001.

[9] T. Lindeberg, *Scale-Space Theory in Computer Vision*. Dordrecht, The Netherlands: Kluwer, 1994.

[10] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

[11] I. C. Sluimer, M. Prokop, and B. van Ginneken, "Towards automated segmentation of the pathological lung in CT," *IEEE Trans. Med. Imag.*, vol. 24, no. 8, pp. 1025–1038, Aug. 2005.

[12] S. Hu, E. A. Hoffman, and J. M. Reinhardt, "Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images," *IEEE Trans. Med. Imag.*, vol. 20, no. 6, pp. 490–498, Jun. 2001.

[13] E. M. van Rikxoort, B. van Ginneken, M. Klik, and M. Prokop, "Supervised enhancement filters: Application to fissure detection in chest CT scans," *IEEE Trans. Med. Imag.*, vol. 27, no. 1, pp. 1–10, Jan. 2008.

[14] S. N. Kalitzin, J. J. Staal, B. M. ter Haar Romeny, and M. A. Viergever, "A computational method for segmenting topological point sets and application to image analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 5, pp. 447–459, May 2001.

[15] J. J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken, "Ridge based vessel segmentation in color image of the retina," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 501–509, Apr. 2004.

[16] D. M. Xu, H. Gietema, H. de Koning, R. Vernhout, K. Nackaerts, M. Prokop, C. Weenink, J. Lammers, H. Groen, M. Oudkerk, and R. van Klaveren, "Nodule management protocol of the NELSON randomised lung cancer screening trial," *Lung Cancer*, vol. 54, no. 2, pp. 177–184, 2006.

[17] M. A. J. Klik, E. M. van Rikxoort, J. F. Peters, H. A. Gietema, M. Prokop, and B. van Ginneken, "Improved classification of pulmonary nodules by automated detection of benign subpleural lymph nodes," in *Int. Symp. Biomed. Imag.*, 2006, pp. 494–497.

[18] M. I. Ahn, A. McWilliams, S. MacDonald, S. Lam, and J. Mayo, "Perifissural opacities (PFO's) detected at chest CT screening for lung cancer," in *Radiological Soc. North Amer.*, 2004, pp. 324–324.