

Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection

Eva M. van Rikxoort^{a)}

Image Sciences Institute, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands

Bartjan de Hoop^{b)}

Department of Radiology, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands

Max A. Viergever^{c)}

Image Sciences Institute, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands

Mathias Prokop^{d)}

Department of Radiology, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands

Bram van Ginneken^{e)}

Image Sciences Institute, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands

(Received 11 December 2008; accepted for publication 15 April 2009; published 9 June 2009)

Lung segmentation is a prerequisite for automated analysis of chest CT scans. Conventional lung segmentation methods rely on large attenuation differences between lung parenchyma and surrounding tissue. These methods fail in scans where dense abnormalities are present, which often occurs in clinical data. Some methods to handle these situations have been proposed, but they are too time consuming or too specialized to be used in clinical practice. In this article, a new hybrid lung segmentation method is presented that automatically detects failures of a conventional algorithm and, when needed, resorts to a more complex algorithm, which is expected to produce better results in abnormal cases. In a large quantitative evaluation on a database of 150 scans from different sources, the hybrid method is shown to perform substantially better than a conventional approach at a relatively low increase in computational cost. © 2009 American Association of Physicists in Medicine. [DOI: [10.1118/1.3147146](https://doi.org/10.1118/1.3147146)]

Key words: segmentation, lung, CT, error detection, automatic

I. INTRODUCTION

The advent of multidetector computed tomography (CT) scanners has made high-resolution, submillimeter isotropic CT imaging the modality of choice for analysis of the lungs. Computerized image analysis of thoracic CT data has become increasingly important over the last decades. For automatic image analysis, lung segmentation is usually an important first step. Accurate lung segmentation allows for the detection and quantification of abnormalities within the lungs. When lung segmentation fails to correctly define the borders of the lungs, findings might be missed or findings outside the lungs might be included in the analysis. Accurate lung segmentation is, for example, important for quantification of emphysema and interstitial lung disease, lung volume estimation, and nodule detection and characterization.

The importance of accurate lung segmentation for a CAD scheme was illustrated by Armato and Sensakovic.¹ They showed that a computer-aided detection system missed 17% of all true nodules due to erroneous lung segmentations. When the lung segmentation paradigm was adapted to the specific task, this improved to only 5% missed nodules. Heussel *et al.*² compared four commercially available programs for quantitative analysis of emphysema. Their results showed that two of the programs were too time consuming to be used in clinical practice because of the necessity of ex-

tensive manual correction of lung segmentation results (59 and 105 min per scan on average). The other two programs required less manual corrections of lung segmentations but still required 26 and 16 min for processing a single scan, owing to manual interaction.

A large amount of research has been devoted to the topic of lung segmentation in 3D CT scans. As a consequence, a wide variety of methods performing automatic segmentation of the lungs is available. Most methods are based on the observation that, for normal lung parenchyma, there is a large difference in attenuation between the lung parenchyma and the surrounding tissue. These conventional methods perform well on scans that do not contain dense abnormalities. However, in the case of dense pulmonary or subpleural abnormalities, these areas are not included in the lung segmentation of these algorithms. This is illustrated in Fig. 1. Other methods are specifically designed to handle such abnormalities, but are too slow or too specialized to be used in clinical practice. An open research area is to investigate the robustness of the various methods.

In practice, there is a trade-off between required computation time and quality of the segmentation results; the results should be as precise as possible and available in as little time as possible. As a result, most available automatic chest CT analysis systems rely on conventional, threshold based,

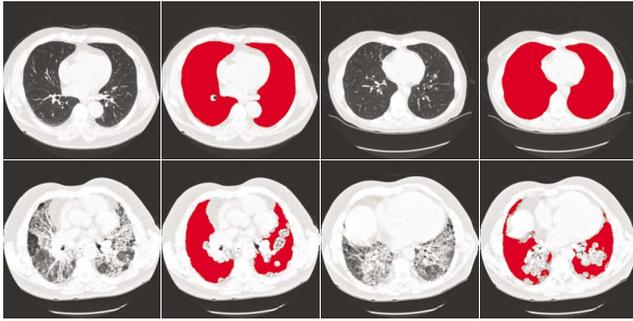


FIG. 1. Illustration of the performance of a conventional lung segmentation method based on gray-value thresholding. The top row shows two slices from two different scans with normal lungs; first the original slice is shown followed by the lung segmentation result overlaid. The conventional method produces a satisfactory segmentation of the lungs. The bottom row shows two slices of two different scans containing pathologic abnormalities. Because of the higher densities of the abnormalities as compared to the density of normal lung parenchyma, the conventional method fails on this type of scans.

segmentation since this is much less time consuming than, for example, a registration-based approach (several minutes vs. several hours) and produces good results in a large number of scans. However, for scans containing (substantial) pathologic abnormalities, such an approach often results in segmentation errors and requires clinicians to manually edit the results. We address this trade-off with a hybrid method that combines a fast conventional approach and a more sophisticated, computationally expensive approach. First, a fast conventional method is applied to each scan. The lung segmentations produced by the conventional method are automatically checked for errors. If there is a reasonable probability that the segmentation result is not accurate, a multiatlas approach is applied. This takes substantially longer but performs better in abnormal cases. A key contribution of this paper is the automatic error detection. Instead of developing a single, compound method for each type of data, we develop a system to analyze the results of the segmentation, which can be used to decide to use a sophisticated method when simpler methods fail.

We applied our hybrid lung segmentation to three databases of 50 scans each. The data originated from different sources, taken under different protocols and with varying types and extents of pathology. In these 150 scans, manual segmentations were performed to allow for quantitative evaluation of the lung segmentation results and of the error detection procedure proposed in this paper.

I.A. Previous work

Normal lungs appear as dark regions in CT data and there is a large difference in attenuation between the lung parenchyma and the surrounding tissue. Most conventional methods to segment the lungs make use of this difference by basing the core of the method on a thresholding approach. Examples of such methods can be found in Refs. 1 and 3–7. Armato and Sensakovic¹ proposed a method that segments the lungs on each slice of the scan separately. First, several

gray-level thresholds were applied to identify the lung regions in each slice and the lungs were separated at the anterior junction. Next, the trachea and the main bronchi were removed from the lung segmentation and holes in the segmentation results due to nodules or vessels were included in the segmentation result. The lung segmentation was applied to scans from two different sources; 38 scans containing annotated lung nodules and 22 scans of patients with a mesothelioma tumor. The results of the lung segmentation itself were not evaluated, only the effect of modifications to the lung segmentation on nodule detection and tumor thickness measurements.

Leader *et al.*³ proposed a similar 2D technique as Armato and Sensakovic.¹ A threshold to identify the lung regions was determined slice by slice. The result was refined by separating the lungs when connected, pruning airway regions included in the result, and filling cavities in the segmentation result. The lung segmentation method was applied to 91 thick slice (5–10 mm) CT scans from 50 emphysema patients and 10 thin slice (2.5 mm) CT scans from a lung cancer screening trial. As a reference standard, the results of a semiautomatic system were used. The results showed that for the thick and thin slice data, the average differences in lung volume between the reference segmentation and the automatic system were 95 and 28 ml, respectively.

Two-dimensional methods are a logical choice when chest CT scans consist of thick slices. When isotropic data are available, it is advantageous to use 3D processing to avoid inconsistencies between slices. Hu *et al.*⁴ extracted the lung regions with an algorithm that operates mainly in 3D. They first applied an optimal thresholding technique followed by a connected component analysis to separate the lungs from other nonbody voxels. Cavities inside the lungs were filled and the trachea and main bronchi were removed from the result. Dynamic programming was applied to separate the left and right lungs on a per slice basis. Finally, the lung boundary was smoothed along the mediastinum. The method was applied to 24 CT scans of eight different patients, manual segmentations were made on every fifth slice for 12 of those scans. The results showed that the automatic lung segmentations closely resembled the manual segmentations, with the root mean square difference between the manual boundary and the automatically determined boundary averaging 0.54 mm.

Sun *et al.*⁵ presented a 3D method in which the lungs were found by region growing based on three criteria: gray value, region homogeneity, and gradient magnitude. After the lungs were extracted, cavities inside the segmented volume were filled using a morphological closing operation. To evaluate the method, it was applied to 20 CT images for which manual tracings of the lung border were available. The results showed that the average volume overlap between the manual delineations and the automatic result was 88.5%.

Brown *et al.*^{6,7} used 3D region growing and morphological operations that were integrated with expert anatomical knowledge in a higher-level set of semantics. In their anatomical model the chest wall, mediastinum, central tracheo-

bronchial tree, and the left and right lung were modeled in terms of attenuation threshold, shape, contiguity, volume, and relative position. The low-level image processing routines were guided by the information from the anatomical model. In Brown *et al.*⁷ the lung segmentation was applied to 104 CT scans of 49 patients. The results were visually judged by a radiologist, no quantitative analysis was performed.

As illustrated in Fig. 1, when high density abnormalities are present in the lungs, conventional methods tend to produce errors. Pu *et al.*,⁸ Prasad *et al.*,⁹ and Sluimer *et al.*¹⁰ developed methods to handle such abnormalities.

The method introduced by Pu *et al.*⁸ was especially designed to include juxtapleural nodules in the lung segmentation result. The image was first smoothed, thresholded at -500 HU, and nonlung regions were flooded slice by slice. Next, the lung borders in each slice were tracked to define the inner lung border. This lung border was used to apply adaptive border marching on each slice to include nodules while minimizing oversegmentation of adjacent regions such as the abdomen and mediastinum. The method was applied to 20 CT scans containing juxtapleural nodules, all nodules in the scans were annotated by consensus reading. The results show that adaptive border marching was able to include all juxtapleural nodules in the lung segmentation result. The typical diameter of all nodules was 4.1 mm, the size of the juxtapleural nodules was not specified. The method is unlikely to work on larger abnormalities, such as those shown in Fig. 1.

Prasad *et al.*⁹ employed the curvature of the ribs in an adaptive thresholding technique to include lung regions that have increased attenuation. As a first step, the ribs were segmented by thresholding and morphologic operations. Next, the lungs were found by varying a threshold until a lung boundary was found that was sufficiently close to the curvature of the ribs. The lung segmentation was applied to 19 scans containing abnormalities due to asthma, scleroderma, and emphysema. The results were compared to a method in which the lung fields were found by thresholding. The method based on rib curvature performed significantly better than the threshold method by including abnormal regions bordering the ribs. However, by lowering the threshold, other nonlung regions not bordering the ribs are included into the segmentation result. In addition, the method depends on the success of the rib segmentation.

Sluimer *et al.*¹⁰ employed a segmentation by registration approach to segment scans containing substantial pathologic abnormalities. A probabilistic atlas was created by registering 15 scans to one reference scan and averaging the result. The segmentation of an abnormal test scan was initiated by elastically registering this probabilistic atlas to the scan and propagating the labels. The resulting segmentation was refined by a voxel classification on the border volume. For a set of ten abnormal scans, manual segmentations were performed on every fourth transversal slice, for the other slices the manual segmentation was interpolated and corrected by the observer. The registration based method outperformed a

region growing method and an interactive region growing method, with overlap fractions with the manual segmentation result of 0.82, 0.39, and 0.79, respectively.

II. METHODS

This section describes the different steps of the hybrid lung segmentation method. The method consists of four steps which comprise three methods. As a first step, the lung fields are segmented with an automatic 3D algorithm, using conventional image processing techniques such as region growing and morphological smoothing (Sec. II A.) This method is fast but likely to produce errors in some abnormal cases. To detect these errors, automatic error detection is applied (Sec. II B). The scans that are likely to contain errors are then segmented by a more sophisticated algorithm, based on multiatlas segmentation (Sec. II C.) This algorithm is likely to perform better for cases containing abnormal pathology but is considerably slower than the conventional algorithm. As a final step, the segmentation results of this algorithm are again checked by the automatic error detection method to determine whether manual interaction is required. Before applying the automatic lung segmentation, all scans were subsampled to axial slices of 256 by 256 and to isotropic voxel sizes in axial direction to reduce required computation times. To obtain the results at the original image resolution, the results were supersampled.

II.A. Conventional lung segmentation

The conventional lung segmentation method that is applied as a first step of the hybrid lung segmentation is comparable to conventional lung segmentation as proposed by Sluimer *et al.*¹⁰ and the algorithm of Hu *et al.*⁴ The algorithm consists of the following steps: (1) Extraction of large airways; (2) Segmentation of lung regions; (3) Separation of the left and right lungs; (4) Smoothing.

II.A.1. Extraction of large airways

Before the lungs are segmented, the trachea and the main parts of the bronchial tree are extracted using a region growing technique. To initialize the region growing, a seed point is automatically identified on the top axial slices of the scan by searching for connected regions with the following properties: an average HU below -950 , a minimum size of 50 mm^2 , a maximum size of 1225 mm^2 , mean x - and y -coordinates not further than 30% of the x - and y -dimensions of the image away from the center of the slice. Starting at the top, the first 25 slices are examined for suitable regions. When several regions are found in a slice, only the region closest to the center of the slice is kept. When all slices have been processed, the region that is closest to the center of its slice is selected. If no suitable region is detected in the top slices, the bottom slices of the scan are inspected to be able to handle cases that were scanned in a reverse direction. The seed point for the region growing process is the voxel with the lowest HU in the found region. From this seed point, the trachea and main stem bronchi are grown

using explosion controlled region growing.¹¹ This approach works as follows. First, region growing is applied from a fixed threshold. Next, the process is iterated, slightly increasing the threshold in each iteration. The growing procedure is stopped when the size of the grown object is increased with (at least) a factor of 2 compared to the size at the previous threshold. Such a volume increase indicates that the algorithm has grown into the lung parenchyma. When this occurs, the tree found by the previous threshold is retained and the increase in threshold is lowered. The process is terminated when no increase in threshold is possible anymore, i.e., increasing the threshold with 1 HU leads to an explosion. For the segmentation of the large airways, the initial threshold was set to -950 HU. In each iteration, the threshold was initially increased with 64 HU. When an explosion occurred, the increase in threshold was divided by 2.

II.A.2. Segmentation of lung regions

After the trachea and main stem bronchi have been determined, the lungs are segmented using region growing. As a seed point for the region growing, the voxel with the lowest HU within the airways is used. From this seed point, the lungs are grown. To determine the upper threshold for the region growing operation, optimal thresholding¹² is applied. After the lung regions have been found, the trachea and bronchi are removed from the result.

II.A.3. Separation of the left and right lungs

Next, the lung region is separated into left and right lungs. When two connected components are found after removal of the trachea and main bronchi, the left and right lungs are identified by their center of gravity. However, in many cases, only one connected component is found because the pleura separating the left and right lung has a gray value included in the region growing. In these cases, the lungs are separated using a dynamic programming approach similar to the technique described by Hu *et al.*⁴ Dynamic programming using the HU of the original image as weights is applied to axial slices in which only one connected component is found. Due to the higher HU of the pleura as compared to that of the lung parenchyma, the maximum cost path corresponds to the pleura separating the lungs. The search area for the dynamic programming is determined in 3D, starting from the center of gravity of the lung region. Let x_{cg} be the x -coordinate of the center of gravity. The search area on axial slices is now defined as all pixels (x, y) for which $x_{cg} - 5 \leq x \leq x_{cg} + 5$. Pixels outside this range are not completely excluded from the dynamic programming procedure, their weight is set to -1000 since it is unlikely that the pleura is located in this area.

II.A.4. Smoothing

As a final step of the lung segmentation method, each lung is smoothed separately. First 3D hole filling is applied to include vessels and other high-density structures that were excluded by the threshold used in region growing, in the

segmentation. Next, morphological closing with a spherical kernel is applied. The diameter of the kernel is set to 2% of the x -dimension of the image.

II.B. Automatic error detection

The goal of the automatic error detection step is to identify scans for which the lung segmentation method may have failed. The automatic error detection consists of two components: comparison with a set of statistics of correct lung segmentations and a shape consistency check.

To develop the error detection method, 335 scans from various sources were used that were not among the test data in this paper. For 320 of those scans, the conventional approach yielded correct lung segmentations (visually confirmed). Those 320 scans were used to determine statistics of correct lung segmentations. The other 15 scans contained erroneous lung segmentations which were used to experiment with measures to detect those errors. Based on experiments using these data, four factors were defined to be important as indicators for success of a segmentation:

- Volume of the left lung;
- Volume of the right lung;
- Ratio between the volumes of the left and right lung;
- Shape consistency of the lung border.

For the first three components, the volumes of the lungs in the 320 scans were determined. Based on these values, the minimum and maximum lung volumes and the ratio between the lung volumes that are considered correct were set to the minimum and maximum values observed in the training data.

Errors in lung segmentation often occur at the borders of the lungs, when the contrast between the lung parenchyma and the surrounding tissue is low due to pathologic abnormalities that show up as dense regions. In normal lung anatomy, the shape of the costal lung surface is convex. When an error occurs at the costal border, the surface is typically not convex anymore. Examples of normal and erroneous lung segmentations are shown in the second column of Fig. 2. Based on this observation, a shape check was developed that tests the convexity of the costal surface in axial slices for each lung separately.

The shape check starts by defining the costal surface of the lungs. For this purpose, the cumulative positions inside the lung segmentation are employed. The cumulative position $(\check{x}, \check{y}, \check{z})$ of a point (x, y, z) inside a lung is defined as follows:

$$\check{x}(x) = \frac{1}{V} \int_0^Z \int_0^Y \int_0^x I(x', y', z') dx' dy' dz',$$

$$\check{y}(y) = \frac{1}{V} \int_0^Z \int_0^y \int_0^X I(x', y', z') dx' dy' dz',$$

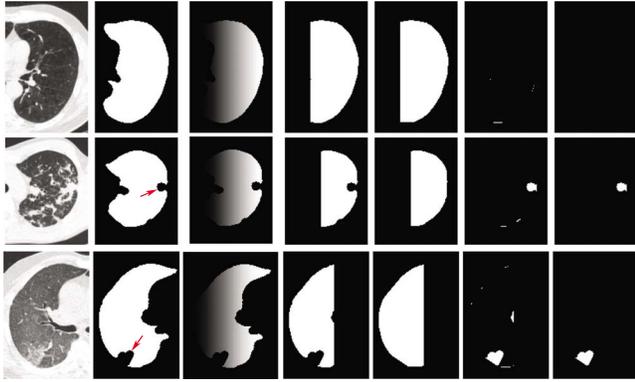


FIG. 2. Illustration of the shape check of the error detection procedure for three slices from three different scans. The first row shows a correct segmentation of a left lung. The second and third rows show erroneous lung segmentations of a left and right lungs, respectively. The errors in the lung segmentations are indicated by arrows. The first column shows the original slice; in the second column the result of the automatic lung segmentation method is shown. The third column shows the cumulative x -positions inside the lung. The third column presents the costal surface of the lung that is taken into consideration. In the fourth column the convex hull is shown; the fifth column shows the difference image between the convex hull and the lung segmentation. Finally, the last column shows the difference between the lung segmentation and the convex hull at the costal surface of the lung after removal of small responses. This difference image is empty for the correct lung segmentation (first row) but not for the incorrect lung segmentations.

$$\check{z}(z) = \frac{1}{V} \int_0^z \int_0^Y \int_0^X I(x', y', z') dx' dy' dz', \quad (1)$$

where V is the total number of voxels in the lung, I is the binary lung image which is 1 inside the lung 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 running from 0 to 1, such that voxels that are at similar spatial locations inside the object in different scans will get approximately the same coordinates. The cumulative x -position inside a lung is illustrated in the third column of Fig. 2. The costal surface in axial slices for the left lung is now defined as those voxels inside the left lung segmentation for which $\check{x} > 0.2$, and for the right lung as those voxels in the segmentation of the right lung for which $\check{x} < 0.8$. The effect of these settings is to leave out the mediastinal region of the lungs for which the convexity constraint does not hold. This is illustrated in the fourth column of Fig. 2.

The convexity is now determined by comparing the costal lung surface in axial slices to the convex hull of this costal lung surface. Examples of the convex hull are shown in the fifth column of Fig. 2. When the difference image between the costal part of the lung and the convex hull is empty, this indicates that the costal lung surface is convex. However, this difference image often contains some small responses at the border (examples are shown in the sixth column of Fig. 2). To remove those responses from the result, a morphological erosion with a spherical kernel of size 7 voxels is applied. Areas remaining in the difference image after the erosion operation likely indicate errors in the segmentation. Ex-

amples of this are shown in the last column of Fig. 2. Since a segmentation error is likely to occur in a larger volume than one axial slice, a scan is marked as probably erroneous if the difference between the convex hull and the lung segmentation occurs in at least three consecutive slices.

The four components of the error detection are combined using a boolean OR operation; if a lung segmentation fails on one of the four tests it is marked as possibly erroneous.

II.C. Multiatlas lung segmentation

The proposed multiatlas segmentation method follows the general scheme of multiatlas segmentation as, for example, described in Rohlfing *et al.*¹³ and Heckemann *et al.*¹⁴ In multiatlas segmentation, a set of training images (atlases) is registered to the image to be segmented. To obtain a segmentation of the target image, the manual labelings of the training images are transformed to the target image using the mapping determined during registration (label propagation). After this process, for each voxel a set of labels is available that are combined by decision fusion to decide on a label for the specific voxel.

Formally, a set of n training images A_i , $i=1, \dots, n$ with corresponding manually segmented (binary) images S_i is available. All n training images are registered to a test image T , resulting in n transformations u_i from A_i to T . u_i is used to transform S_i to T . We write the transformed image as $u_i(S_i)$. As a result, for each voxel in T , n opinions exist about its label. The labels are combined by averaging all opinions for each voxel. After n registered training images, S is defined as

$$S = \frac{1}{n} \sum_{i=1}^n u_i(S_i). \quad (2)$$

To obtain a binary segmentation S is thresholded at 0.5.

After averaging the labels, each lung is smoothed separately using the same smoothing as described for the conventional lung segmentation. Morphological closing with a spherical kernel with a diameter of 2% of the x -dimension of the image is applied.

II.C.1. Registration

In registration, an atlas image is transformed to the target image. In this paper, ELASTIX (ELASTIX can be downloaded from <http://elastix.isi.uu.nl>) version 3.9 was used for registration. This framework formulates the registration problem as an optimization problem in which the similarity between the target and atlas image is maximized. The framework is largely based on techniques described in Thévenaz *et al.*,¹⁵ Rueckert *et al.*,¹⁶ and Mattes *et al.*¹⁷ Many parameters in the algorithm can be chosen. For the cost function the negative mutual information was used, which is implemented in ELASTIX according to Thévenaz *et al.*¹⁵ First, images were roughly aligned with an affine transformation. After that, a nonrigid registration modeled by B-splines was performed. For the optimization of the cost function, an iterative stochastic gradient descent optimizer¹⁸ was used. In each iteration a step toward the minimum is taken, and this direction is

TABLE I. Specifications of the properties of the scans in the different data sets. The ranges of the values for exposure, tube voltage, slice spacing, and pixel size are provided.

Database	Exposure (mAs) ^a	Tube voltage (kVp)	Slice spacing (mm)	Slice thickness (mm)	Pixel size (mm)	Scanners (number)
NELSON	30	120–140	0.7	1.0	0.75	Phillips Brilliance-16 (50)
LIDC	ambiguous	120–140	0.75–3	1.5–3	0.531–0.836	GE Medical Systems LightSpeed QX/i (16), GE Medical Systems LightSpeed Ultra (4), GE Medical Systems LightSpeed16 (1), GE Medical Systems LightSpeed Pro 1 (1), Toshiba Aquilion (10), Siemens Sensation 16 (9), Siemens Emotion 6 (3), Philips Brilliance 40 (1), Philips Brilliance 16P (5)
ILD	53–120/ ambiguous	120–140	0.8	0.8	0.49–0.75	Philips Mx8000 IDT (50)

^aFor all scans in the LIDC data base and a number of scans in the ILD database no exposure values were provided in the DICOM header.

based on the derivative of the cost function to the transformation parameters. To speed up the registration, the derivative is calculated on a small randomly chosen subset of samples in each iteration. To avoid local minima, a multi-resolution strategy was taken.

Registration parameters were determined in a set of pilot experiments. For the affine registration four resolutions were used, in each of which 512 iterations of the stochastic gradient descent optimizer were performed. For the nonrigid B-spline registration five resolutions were used. The B-spline grid spacings used in these resolutions were 64, 64, 32, 16, and 8 voxels, respectively. The optimizer performed 1024 iterations in each resolution. For both affine and nonrigid registration 32 histogram bins were used. With these settings, one registration takes about 15 min on a standard high-end PC. Specifications about the atlases used for the lung segmentation are provided in Sec. III A.

III. DATA

For the evaluation of this study, three different groups of data were selected.

- (1) 50 scans were randomly selected from a lung cancer screening trial (NELSON).¹⁹
- (2) 50 scans were randomly selected from the publicly available repository of the lung image database consortium (<http://nci.nih.gov>) (LIDC).²⁰
- (3) 50 scans were randomly selected from a database containing patients with confirmed interstitial lung disease (ILD).

The rationale behind using data from those three different sources is twofold. First, by selecting data from different sources we make sure that a variety of scanning protocols, scanner types, and image resolutions is represented. Second, the amount and extent of pathologic abnormalities vary substantially between the different groups. The subjects participating in the NELSON study are asymptomatic and most

scans do not contain substantial pathologic abnormalities, except for emphysema and pulmonary nodules. The aim of the lung image database consortium is to develop an image database as a research resource for the development, training, and evaluation of CAD methods using CT imaging.²⁰ The data in the LIDC database were collected from five academic hospitals in the USA. No requirements with regard to scanner pitch, exposure, tube voltage, or reconstruction algorithm were imposed. As a result, the LIDC database contains CT scans from a range of manufacturers and models, acquired with different scan protocols. The database contains CT scans from both diagnostic and screening studies, which leads to a variety of abnormalities being present. All scans in the ILD database are from patients with confirmed interstitial lung disease. The data were collected from clinical practice in several hospitals in the Netherlands. Types of interstitial lung disease include sarcoidosis, idiopathic interstitial pneumonia, and various immune and autoimmune disorders. The severity of the abnormalities varies substantially between patients. Table I provides the details for the different data sets. All scans were obtained at full inspiration, one scan in the ILD database and ten scans in the LIDC database were obtained with contrast material.

III.A. Atlases

As described in Sec. II B, a set of images with lung segmentations is needed to perform the multiatlas segmentation method. For each test scan, eight atlases were used during the multiatlas segmentation. Two different sets of atlases were used to segment the data from the different sources. For test scans from the NELSON and LIDC databases, eight scans from the NELSON study were used as atlases. These eight scans were not among the 50 NELSON scans used for evaluation in this paper. For the ILD database, eight scans from the ILD database, not among the 50 scans used in this paper, were selected. In both sets of atlases, the lungs were segmented by a conventional approach. All segmentations

were visually checked to be correct. The rationale behind the different sets of atlases for the ILD database is that the variety in protocol and abnormalities that is present in the ILD database is better represented by this set of atlases than by atlases selected from the NELSON database.

IV. EXPERIMENTS

The hybrid lung segmentation method was applied to all 150 scans. To allow for a quantitative evaluation, the lungs in all 150 scans were manually segmented at the original resolution. In this section, the protocol for manual segmentation, the evaluation measures, and the experiments will be described.

IV.A. Manual segmentations

To allow for a quantitative evaluation of the hybrid lung segmentation method, three human observers manually segmented the lungs in a set of slices in each direction, in all 150 scans. The slices that were manually segmented were equally spaced between the start and end points of the lungs. To determine the slices that had to be manually segmented, a bounding box around the lungs was manually set by one human observer for each scan. This bounding box was used to determine in each direction (sagittal, coronal, axial), between three and five slices to be segmented. This resulted in 9–15 slices per lung that were manually segmented. In total 1800 slices were manually segmented.

Manual delineation of the lung border was performed by clicking points on the border. These points were automatically connected by straight line segments. Observers had the option to change these line segments by adding or moving points. The observers were able to scroll through the scan in all directions while segmenting a slice. For manual delineation, a standard lung window level was used (window level of -600 , width of 1400).

A difficult part in segmenting the lung fields is the hilar region. Ideally the segmentation of the lung volume would intersect the bronchial and pulmonary vessel tree at the pleural fold. However, since the pleural fold is not visible in 3D thoracic CT data, we developed a protocol to improve inter-observer agreement in the hilar region. The protocol used is given in Appendix A.

The lungs in all 150 scans were manually segmented by three observers, which allows for quantification of the inter-observer variability. The manual segmentations were made by medical students who were trained by a radiologist. One observer segmented all 150 scans, those segmentations served as the reference standard. The other two observers together segmented all 150 scans, those segmentations were used to estimate interobserver variability. To ensure that the observers were sufficiently skilled at segmenting the lungs, a training procedure was set up. First, a session was organized with a radiologist in which the main anatomy was reviewed and examples were shown. Next, a set of five training scans was provided to each observer for which a reference was known. The observers were asked to segment those scans as a training session. The segmentations as produced by each

observer were compared to the reference. When the observers' performance was satisfactory they were allowed to start segmenting the data used in this paper.

On average, manually segmenting one lung in one slice took 1.2 min. So, segmenting both lungs in all 150 scans took 72 h. Since every scan was manually segmented twice, the total time spent on manually delineating the data was 144 h.

IV.B. Evaluation of lung segmentation

The performance of the lung segmentation method was quantitatively evaluated by comparing it to the manual reference standard. In addition, the performance of the hybrid lung segmentation method was compared to the performance when only the conventional method was used. The evaluation was performed using the following measures: volumetric overlap, Hausdorff distance, and mean absolute surface distance.²¹

The volumetric overlap fraction is a volumetric measure that compares the relative overlap of two binary segmentations. For two volumes V_{segm} and V_{manual} the volumetric overlap fraction is defined as the volume of their intersection divided by the volume of their union,

$$O(V_{\text{segm}}, V_{\text{manual}}) = \frac{|V_{\text{segm}} \cap V_{\text{manual}}|}{|V_{\text{segm}} \cup V_{\text{manual}}|}. \quad (3)$$

The more equal two volumes are, the closer O is to 1. O is a volumetric measure and it therefore depends on the size of the object. Since a lung is a large object, small local errors at the boundary are not captured in the overlap measure. For this reason the Hausdorff distance is also computed.

The Hausdorff distance (or maximum surface distance) defines the largest difference between two contours. Given two surfaces S_{segm} and S_{manual} let $d_{\min}(p, S_{\text{manual}})$ be the minimum distance of a point p on the surface S_{segm} to the surface S_{manual} . The largest minimum distance from all points p on the surface S_{segm} to the surface S_{manual} is then defined as

$$h(S_{\text{segm}}, S_{\text{manual}}) = \max_{p \in S_{\text{segm}}} (d_{\min}(p, S_{\text{manual}})). \quad (4)$$

The Hausdorff distance is now defined as a symmetric extension,

$$H(S_{\text{segm}}, S_{\text{manual}}) = \max(h(S_{\text{segm}}, S_{\text{manual}}), h(S_{\text{manual}}, S_{\text{segm}})). \quad (5)$$

The Hausdorff distance is very sensitive to outliers and therefore could give a distorted impression of the quality when used on its own. However, in combination with other surface based measures it is informative. Therefore, as a final measure the mean absolute surface distance was used.

The mean absolute surface distance is a measure for how much on average two surfaces differ. Let $\bar{d}_{\min}(S_{\text{segm}}, S_{\text{manual}})$ be the average minimum distance from all points on the surface S_{segm} to the surface S_{manual} , then the mean absolute surface distance between S_{segm} and S_{manual} is defined as

TABLE II. Results of the automatic system and of the second human observer as compared to the reference segmentation in terms of volumetric overlap (O), Hausdorff distance in mm (H), and mean absolute surface distance in mm (M). The mean and standard deviation (SD) are given for each measure.

Measure	Second observer		Hybrid		Conventional ^a	
	Mean	SD	Mean	SD	Mean	SD
Total						
O	0.96	0.04	0.95	0.04	0.93	0.14
H	22.84	14.31	23.55	12.70	25.79	17.48
M	1.75	4.28	1.61	3.13	1.64	3.13
NELSON						
O	0.96	0.05	0.96	0.02	0.94	0.14
H	20.23	17.83	18.85	11.07	18.95	11.00
M	1.84	5.00	1.06	0.48	1.07	0.49
LIDC						
O	0.96	0.04	0.95	0.05	0.91	0.19
H	23.63	11.70	23.29	11.22	24.95	11.60
M	1.50	4.20	1.69	4.09	1.72	4.11
ILD						
O	0.95	0.03	0.95	0.03	0.94	0.03
H	25.11	12.35	28.75	14.13	35.24	24.40
M	1.91	3.64	2.08	3.49	2.12	3.41

^aFor 3 out of the 150 scans the conventional method was not able to identify any lung volume, leading to an empty segmentation. In those cases M and H are undefined and therefore left out of the results. The true performance of the conventional method on H and M in total and for the NELSON and LIDC database are thus worse than indicated in this table.

$$M(S_{\text{segm}}, S_{\text{manual}}) = (\bar{d}_{\min}(S_{\text{segm}}, S_{\text{manual}}) + \bar{d}_{\min}(S_{\text{manual}}, S_{\text{segm}}))/2, \quad (6)$$

O , H , and M were computed at full resolution. To determine which voxels in the segmentation were on the surface, 26-connectivity was used.

IV.C. Evaluation of error detection

The automatic error detection was evaluated using the measure described in the previous section; the values for volumetric overlap O , Hausdorff distance H , and mean absolute surface distance M for the scans marked as erroneous were compared to the values for the scans marked as correct. In addition, an observer study was set up. The lung segmentations that were detected as erroneous were shown to a radiologist. The radiologist was not informed that the lung segmentations were marked as wrong by our system. For each scan, the radiologist was asked to indicate if there was an error in the lung segmentation.

V. RESULTS

V.A. Evaluation of lung segmentation

The hybrid lung segmentation approach was applied to all 150 scans. After the conventional method 22 errors were detected, those 22 scans were segmented with the multiatlas method. Table II lists the results of the hybrid and conventional segmentation methods for each measure, in total as

well as for each database separately. The results of the second observer as compared to the reference segmentation are also provided. It can be seen that for the total database of 150 scans, O , H , and M are better for the hybrid method than for the conventional method. Note that H and M should be small for good segmentations, whereas O should be large. In Fig. 3, the results of the hybrid method for all scans in which an error was detected are provided.

The conventional method takes on average 55 s per scan, the error detection takes on average 4 s per scan, and a multiatlas segmentation takes on average 120 min. Given this, the hybrid lung segmentation takes on average 18.6 min (22×120 min and 128×59 sec) per scan for the complete database of 150 scans.

V.B. Evaluation of error detection

In total, 22 errors were automatically detected in the conventional method. After the registration based segmentation had been applied to those scans, four errors were still detected. Table III provides the division of the errors over the different data sets as well as a specification of the error found. Figure 3 shows a slice of the hybrid lung segmentation for each of the 22 errors detected after the conventional method in which the segmentation error is most obvious. Both the segmentation results of the conventional method and that of the hybrid method are provided. From Fig. 3 different reasons why the conventional lung segmentation method fails can be identified.

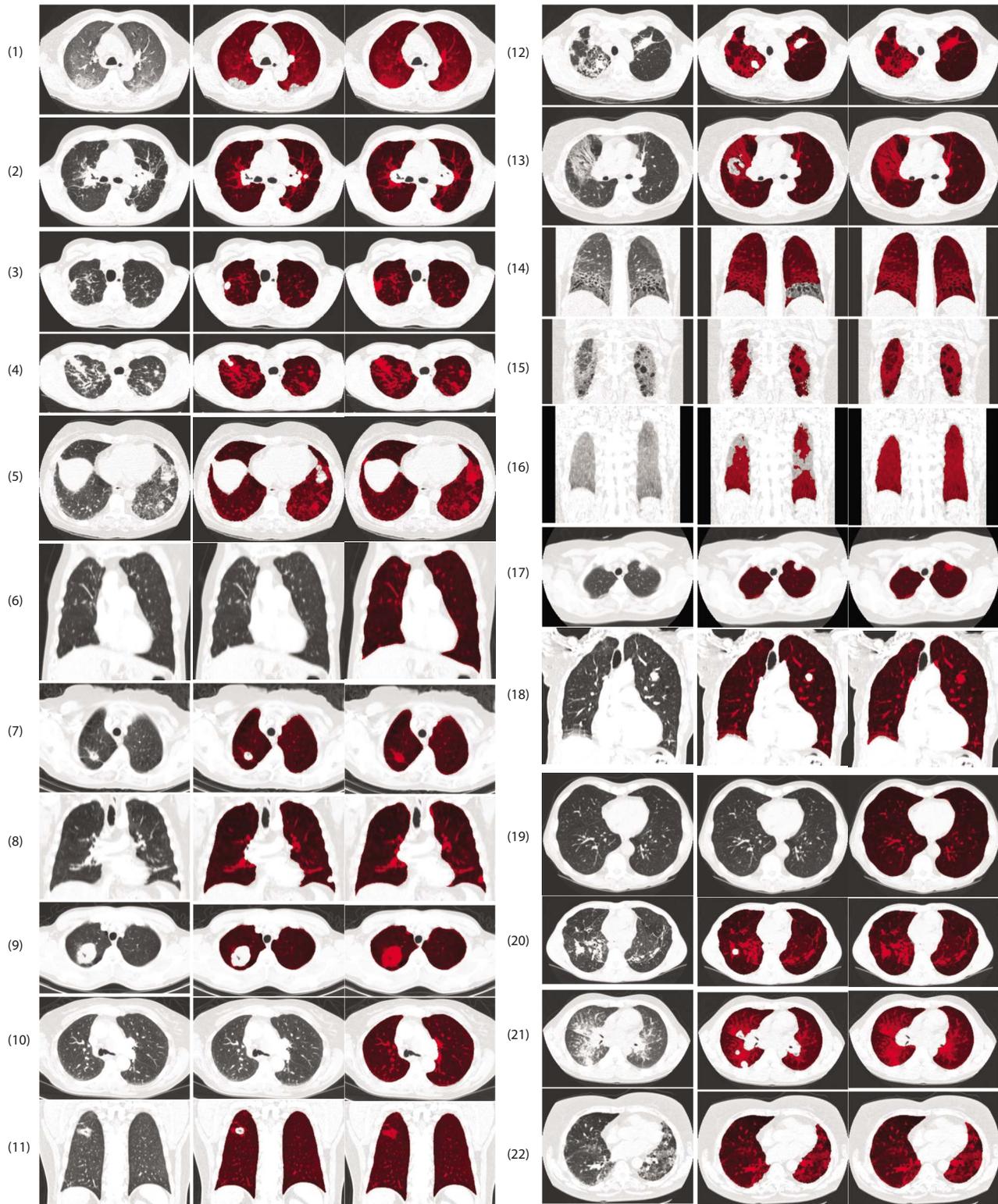


FIG. 3. Results of the hybrid lung segmentation for the 22 scans in which an error was detected after the conventional lung segmentation method. The first column shows the original slice, the second column shows the conventional lung segmentation, and the last column shows the result of the hybrid method.

- Abnormalities not are included in the segmentation: (1–5, 7–9, 11–18, 20–21).
- One or both lungs were not found (6, 10, 19).

For scan 22 in Fig. 3 an error was detected based on too low lung volume; however, the segmentation was correct.

When one or both of the lungs are not found, this is due to the airway segmentation that is used as a starting point for the lung segmentation; the automatic initialization of the airway segmentation failed. For two scans, no suitable seed point was detected at all; for the other scan, the seed point

TABLE III. Number and type of errors detected in the hybrid lung segmentation method per database. The 22 errors found after applying the conventional method are specified in the second and third columns. Specifications of the four errors detected after applying the hybrid method are provided in the last two columns.

Database	Conventional method		Hybrid method	
	No. of errors	Type of error (No.)	No. of errors	Type of error (No.)
Nelson	1	No lungs found (1)	0	...
LIDC	9	Shape error (6) Lung volume too low (1) No lungs found (2)	1	Lung volume too low (1)
ILD	12	Shape error (7) Lung volume too low (5)	3	Lung volume too low (3)

was erroneously placed in the bowels, which led to a seed point from which the lungs could not be segmented.

Table IV gives the results in terms of O , H , and M for cases in which errors were detected. It can be seen that the hybrid lung segmentation method performs substantially better than the conventional method for those 22 scans. When the 22 scans for which the conventional method failed are removed from the 150 scans, the performances of the conventional method in terms of O , H , and M are 0.96 ± 0.03 , 23.58 ± 12.87 , and 1.57 ± 3.22 mm, respectively.

Since errors are often local, an additional evaluation was applied in which O , H , and M were calculated only for the slice in which the difference between the hybrid method and the conventional method was largest. Note that only the slices for which manual segmentations were made could be used, so it is not guaranteed that the slice where the error occurred is available if the error was (very) small. The results are listed in Table V. For this evaluation, the difference between the hybrid method and the conventional method is more pronounced.

The observer study as described in Sec. IV C revealed that out of the 22 detected errors, one segmentation was correct (this corresponds to image 22 in Fig. 3). The lung segmentation was marked as erroneous based on a low lung volume. After the multiatlas approach, four scans were marked as possibly erroneous based on a low lung volume by the error detection. The observer study showed that those four segmentations were actually all correct lung segmentations. These four scans contained the scan that was erroneously marked as wrong after the first segmentation.

VI. DISCUSSION

Segmentation of structures in medical images is challenging due to, among other things, anatomical differences, abnormalities, image noise, and differences in acquisition parameters. Segmentation of the lungs in thoracic CT scans is an important prerequisite for automatic analysis of chest CT scans. Each available chest CT analysis system incorporates an automatic lung segmentation. However, the lung segmentation methods incorporated in most automatic analysis systems are conventional methods based on thresholding and region growing since these techniques are generally fast. The lung segmentations produced by such systems are often erroneous for clinical scans, which leads to incorrect analysis of the data. The discrepancy between the results reported in the literature and the actual results in a clinical setting are partly caused by the fact that many algorithms have only been tested on very select patient populations that do not necessarily represent clinical practice.

VI.A. Evaluation

To test the robustness of a segmentation method, it is important to apply it to a large database of scans containing a variety of properties. Since manually delineating lungs in a large set of scans is tedious and tiring work, acquiring data for quantitative evaluation is often a limiting factor. In this paper, we chose an approach in which a set of 9–15 slices were segmented per scan. The slices were chosen in such a way that a substantial part of the lungs was covered in the manual segmentation. This approach was preferred over

TABLE IV. Results of the automatic system and of the second human observer as compared to the reference on the 22 scans in which errors were detected. Mean and standard deviation (SD) are given for volumetric overlap (O), Hausdorff distance (H), and mean absolute surface distance (M).

Measure	Second observer		Hybrid method		Conventional method	
	Mean	SD	Mean	SD	Mean	SD
O	0.96	0.02	0.94	0.03	0.81	0.28
H	18.41	11.31	23.53	11.97	41.92	10.60
M	1.55	2.31	1.90	2.70	2.15	2.43

manually segmenting complete lungs, since in this way, in the same time, manual segmentations for many more scans could be obtained. In total, it took 144 h to segment the slices in all 150 scans twice.

The 150 scans used for evaluation contained a large variety in acquisition parameters and abnormalities. The segmentations were evaluated using three different methods that each evaluates a different aspect of the segmentation result. The results as shown in Tables II, IV, and V show the usefulness of both the variety in data that was used and the different evaluation measures. As expected, least errors were found in the scans from the NELSON database, which contains scans from asymptomatic subjects. Most errors were detected in the ILD database, containing scans from patients with confirmed interstitial lung disease. The evaluation clearly shows that using the hybrid method becomes more advantageous as the data get more challenging to segment; for the ILD database, the difference between the results of the conventional method and those of the hybrid method is largest. The results for the LIDC database show that the method is robust to variations in scanner types and acquisition protocols.

VI.B. Segmentation accuracy

Figures 1 and 3 illustrate that conventional lung segmentation methods do not produce an accurate segmentation of the lungs in the presence of high density abnormalities. The conventional lung segmentation that was used in this paper failed in 19 of such cases, whereas it accurately segmented 128 scans in which no such abnormalities were present (see Table II). When no abnormalities were present, the conventional method was often more accurate than the multiatlas method and even outperformed the manual segmentation. Especially at the lung borders, where the contrast between the lung parenchyma and the surrounding tissue is high, thresholding produces excellent results. This is illustrated in Fig. 4, where it can be seen that the conventional method follows the lung border more accurately than both the manual delineation and the multiatlas method. The difference with the manual segmentation is caused by the protocol; the lung border was segmented by clicking points on the lung border, which inevitably leads to slight inaccuracies. In ad-

dition, the window level used during manual contouring may influence the position of the manually drawn border.

Although the manual segmentations and the automatic segmentations visually look similar for most scans, the evaluation measures are slightly better for a second observer than for the automatic method. Especially at the mediastinum, the automatic method and the manual segmentations tend to be different because of the low contrast. The manual segmentations are performed using a set of rules, see Appendix A, whereas the automatic segmentations are either based on a threshold or on registrations. Since both manual segmentations were performed according to the same rules, they are more consistent.

VI.C. Hybrid segmentation approach

The hybrid approach applied in this paper consists of two segmentation methods, connected by an error detection method. By applying the hybrid segmentation approach to a large database of scans from different sources and containing a variety of abnormalities, the hybrid method was tested for robustness. Table II shows that the hybrid approach performs better than the conventional approach for the total database of 150 scans.

Another advantage of the hybrid approach is the relatively low segmentation time. If only the multiatlas approach would have been applied, similar performance would have been obtained but the average segmentation time would be 120 min per scan. By only applying the conventional method, the average segmentation time would have been 55 s per scan. However, 22 scans would have required subsequent manual interaction, which would have led to a much higher total segmentation time per scan. By applying the hybrid method, the average segmentation time was 18.6 min per scan for the total database of 150 scans. For the NELSON database, in which few abnormalities are present, the average segmentation time was 3.4 min, since only one conventional lung segmentation failed. For the ILD database, in which most errors were detected, the average segmentation time was 30 min. All computation times reported in this paper were done on a single core computer. Both algorithms, the conventional method and the multiatlas method, are easy to parallelize, so on a quadcore machine a speed-up factor of close to 4 can be achieved.

TABLE V. Results of the automatic system and of the second human observer as compared to the reference segmentation on one slice of each of the 22 scans in which errors were detected. The slice was selected from the slices for which a manual segmentation was available, as the slice for which the hybrid method and the conventional method differed most on average over the three measures [volumetric overlap (O), Hausdorff distance (H), and mean absolute surface distance (M)]. Mean and standard deviation (SD) are given for each measure.

Measure	Second observer		Hybrid method		Conventional method	
	Mean	SD	Mean	SD	Mean	SD
O	0.96	0.02	0.95	0.03	0.81	0.33
H	6.85	9.23	9.22	7.10	41.57	21.22
M	0.67	0.82	0.49	0.37	2.96	1.43

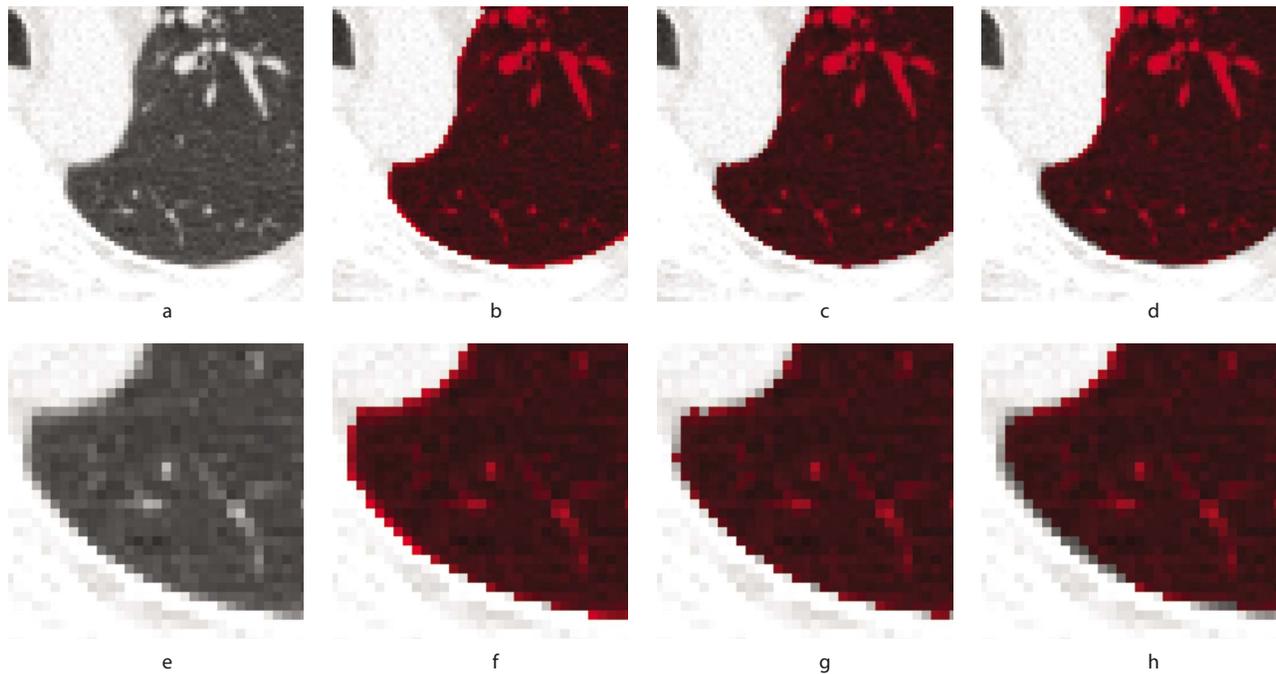


FIG. 4. Illustration of the difference between the manual segmentation, the conventional method, and the multiatlas segmentation at the lung border. The first row shows part of a slice of an image, with the lung segmentations overlaid. The second row shows a more zoomed in version of the same slice. In (a) and (e), the original image is shown; (b) and (f) provide the manual segmentation. In (c) and (g), the segmentation of the conventional method is shown, (d) and (h) show the multiatlas segmentation. It can be seen that the conventional method follows the lung border more precise than both the multiatlas method and the manual tracing.

VI.D. Error detection

The key element of the hybrid segmentation method is automatic error detection. Good error detection facilitates an optimal combination of different methods. The error detection method as developed for this paper is fast (4 s per scan) and the evaluation described in Sec. V shows that the method performs well and is able to detect both large and small errors. The small errors do not substantially deteriorate the overall segmentation result much as captured by the global measures O , H , and M , but they are important in CAD systems; if the abnormality is not captured in the lung segmentation, it will not be picked up by an automatic detection system.

To capture small local errors, the 22 scans in which an error was detected were evaluated on the slice in which the hybrid method and the conventional method differed most. The results as presented in Table V show that the Hausdorff distance H is very high for the conventional method on those slices. Since H measures the largest difference between the automatic segmentation and the reference standard, this means that, although the larger part of the segmentation is probably correct given the values for O and M for the complete scans, there is a position in a slice for which the conventional method is far from the manual contour. This results shows the importance of error detection; while global evaluation measures like O , H , and M provide reasonable values for many scans, the slice-based evaluation shows that there was a (substantial) local error which was correctly detected.

Since the data in the LIDC data base contain most variation in scanner type, protocol, slice spacing, and slice thickness, we took a closer look at the nine errors detected in the LIDC scans after the conventional method. The results revealed that the errors were not exclusive for a certain slice spacing, slice thickness, or scanner type: slice spacing ranged from 0.75 to 3 mm, slice thickness ranged from 1.5 to 3 mm, and the nine scans were made by six different scanners from four different manufacturers.

The errors detected in the ILD database were either due to too low a lung volume or to a shape error detected at the costal surface of the lungs. A closer look at the scans reveals that from the five scans detected as erroneous based on too low lung volume, four scans were incorrectly segmented because of extensive pathologic abnormalities not included in the segmentation. This is for example illustrated by case 20 in Fig. 3. For the other scan in which a low lung volume was detected the segmentation was correct.

VI.E. Improvements

By inspecting the results of the conventional method for the 150 scans used in this paper, several possible improvements for the conventional method can be identified. First of all, for three scans no lungs were found owing to a wrong initialization of the bronchial tree segmentation. This could be improved by, for example, using a method that is forced to grow a tubular structure after the seed point has been identified. Although this might solve the errors as detected in

the three scans in this paper, it is difficult to define a robust seed point detector for the airway tree because of difference in scan protocols, anatomy, and abnormalities. The second improvement could be to incorporate the method as proposed by Pu *et al.*,⁸ in which small errors at the lung border are included in the lung segmentation (see Sec. I A). This would, for example, fix the error as detected in case 4 of Fig. 3. However, it would probably fail for the type of abnormality shown in case 16 of Fig. 3. Finally, the method as proposed by Prasad *et al.*⁹ would be able to handle certain errors localized at the outer part of the lungs (see Sec. I A). However, as indicated by Prasad *et al.*⁹ themselves, this method is likely to introduce errors in other areas in the scan when the threshold is lowered too much.

As was indicated in the discussion about segmentation accuracy, multiatlas segmentation is robust but tends to deviate slightly from the correct lung border. In most cases this deviation is small. However, in cases where the deviation is more substantial the segmentation could be improved by voxel classification of the border voxels, as proposed by Sluimer *et al.*¹⁰

In the current setup of the multiatlas method two different sets of atlases are used, one for the NELSON and LIDC database and one for the ILD database. This is not optimal as concerns robust applicability of our segmentation approach since it implies that for different data different atlases should be used. The use of different atlas sets could be solved by applying an atlas selection scheme. One large set of atlases could be used for all types of data, from which an appropriate subset of atlases would automatically be selected for each test scan. Such a scheme is for example described in Aljabar *et al.*²²

The setup of the hybrid method as presented in this paper could be improved by adding more submethods and by specializing the error detection. In the current approach, two algorithms are combined using one general error detection method that indicates whether a segmentation is correct or not. By specializing the error detection to indicate the type and severity of the error as well as the location, a system could be designed that starts with a conventional computationally cheap method (like the method used in this paper). When an error is detected, depending on the type of error, another method is started. Note that such a design would increase the overall complexity of the method.

VI.F. Challenges

The hybrid approach as presented in this paper is a step toward making automatic lung segmentation more robust and reliable. By combining the strengths of different methods, a variety of segmentation errors can be solved. However, there are still scans that are challenging to segment for every automatic segmentation method currently available, e.g., scans containing pleural effusions. It is often very difficult for a radiologist to define the correct lung border in such scans. For automatic methods it is even harder owing to the different appearances of the lung border. In addition, there are

cases in which both the density and the shape of the lung are altered by abnormalities. In such cases, both a shape based and a conventional method are likely to fail.

By acknowledging that certain scans might be too difficult to segment automatically, error detection becomes important in a different way. It would be used to indicate to the radiologist which scans need manual editing. This would be a practically useful addition to existing clinical segmentation methods.

VII. CONCLUSION

Lung segmentation is a prerequisite for automated analysis of chest CT scans. Conventional lung segmentation methods rely on large attenuation differences between lung parenchyma and surrounding tissue. These methods fail in scans where dense abnormalities are present, which holds for a large percentage of clinical scans. In this paper, a new hybrid lung segmentation method was presented that automatically detects failures of a conventional algorithm and, when needed, resorts to more a complex algorithm, which produces better results in abnormal cases. A quantitative evaluation on a large database collected from different sources showed that the hybrid method performs better than a conventional approach at a relatively low increase in computational cost.

APPENDIX A: PROTOCOL FOR MANUAL LUNG SEGMENTATION

The objective is to segment the lung boundaries manually. At the mediastinum the lung boundary is hard to determine, therefore a protocol was setup to make the segmentation more consistent.

Although anatomically it would be preferable to let the segmentation of the lung volume intersect the bronchial and pulmonary vessel trees at the pleural fold, this ring cannot be reliably located in the 3D CT volume since it is not directly visible.

The following rules indicate where to let the lung segmentation intersect the pulmonary vessels and bronchi in the hilar region. Parts of the bronchial and pulmonary vessel tree distal from the intersections mentioned below are considered to be included in the lung volume. Numbers between brackets refer to bronchial branches as indicated in Fig. 9.1 on p. 282 in Prokop and Galanski.²³

Right lung:

- Upper lobe: immediately before the first division of the right upper lobe bronchus.
- Middle lobe: immediately before the first division of the right middle lobe bronchus (division into 4/5).
- Lower lobe: immediately before the division of the right lower lobe bronchus into the basal segmental bronchi (division into 7/8/9/10). Note that the apical or superior segmental bronchus of the right lower lobe (6) is to be included in the lung volume.

Left lung:

- Upper lobe: immediately before the first division of the left upper lobe bronchus.
- Lower lobe/lingula: immediately before the first division of the left lower lobe lingular bronchus (division into 4/5).
- Lower lobe: immediately before the division of the left lower lobe bronchus into the basal segmental bronchi (division into 7/8/9/10). Note that the apical or superior segmental bronchus of the left lower lobe (6) is to be included in the lung volume.

^{a)}Electronic mail: eva@isi.uu.nl

^{b)}Electronic mail: b.j.dehoop@umcutrecht.nl

^{c)}Electronic mail: max@isi.uu.nl

^{d)}Electronic mail: m.prokop@umcutrecht.nl

^{e)}Electronic mail: bram@isi.uu.nl

¹S. G. Armato and W. F. Sensakovic, "Automated lung segmentation for thoracic CT: Impact on computer-aided diagnosis," *Acad. Radiol.* **11**(9), 1011–1021 (2004).

²C. P. Heussel, T. Achenbach, C. Buschsieweke, J. Kuhnigk, O. Weinheimer, G. Hammer, C. Düber, and H.-U. Kauczor, "Quantification of pulmonary emphysema in multislice-CT using different software tools," *Röfo* **178**(10), 987–998 (2006).

³J. K. Leader, B. Zheng, R. M. Rogers, F. C. Scierba, A. Perez, B. E. Chapman, S. Patel, C. R. Fuhrman, and D. Gur, "Automated lung segmentation in X-ray computed tomography: Development and evaluation of a heuristic threshold-based scheme," *Acad. Radiol.* **10**(11), 1224–1236 (2003).

⁴S. 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**(6), 490–498 (2001).

⁵X. Sun, H. Zhang, and H. Duan, "3D computerized segmentation of lung volume with computed tomography," *Acad. Radiol.* **13**(6), 670–677 (2006).

⁶M. S. Brown, M. F. McNitt-Gray, N. J. Mankovich, J. G. Goldin, J. Hiller, L. S. Wilson, and D. R. Aberle, "Method for segmenting chest CT image data using an anatomical model: Preliminary results," *IEEE Trans. Med. Imaging* **16**(6), 828–839 (1997).

⁷M. S. Brown, J. G. Goldin, M. F. McNitt-Gray, L. E. Greaser, A. Sapra, K. T. Li, J. W. Sayre, K. Martin, and D. R. Aberle, "Knowledge-based segmentation of thoracic computed tomography images for assessment of split lung function," *Med. Phys.* **27**(3), 592–598 (2000).

⁸J. Pu, J. Roos, C. A. Yi, S. Napel, G. D. Rubin, and D. S. Paik, "Adaptive border marching algorithm: Automatic lung segmentation on chest CT images," *Atmos. Technol.* **32**(6), 452–462 (2008).

⁹M. N. Prasad, M. S. Brown, S. Ahmad, F. Abtin, J. Allen, I. da Costa, H. J. Kim, M. F. McNitt-Gray, and J. G. Goldin, "Automatic segmentation of

lung parenchyma in the presence of diseases based on curvature of ribs," *Acad. Radiol.* **15**(9), 1173–1180 (2008).

¹⁰I. 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).

¹¹K. Mori, J. Hasegawa, J. Toriwaki, H. Anno, and K. Katada, "Recognition of bronchus in three-dimensional X-ray CT images with application to virtualized bronchoscopy system," *Proceedings of the International Conference on Pattern Recognition*, 1996 (unpublished), Vol. 3, pp. 528–532.

¹²N. Otsu, "A threshold selection method from gray level histograms," *IEEE Trans. Syst. Man Cybern.* **9**(1), 62–66 (1979).

¹³T. Rohlfing, R. Brandt, R. Menzel, and C. R. Maurer, Jr., "Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains," *Neuroimage* **21**(4), 1428–1442 (2004).

¹⁴R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers, "Automatic anatomical brain MRI segmentation combining label propagation and decision fusion," *Neuroimage* **33**(1), 115–26 (2006).

¹⁵P. Thévenaz, T. Blu, and M. Unser, *Handbook of Medical Imaging, Processing and Analysis* (Academic, San Diego, CA, 2000), pp. 393–420.

¹⁶D. 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).

¹⁷D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank, "PET-CT image registration in the chest using free-form deformations," *IEEE Trans. Med. Imaging* **22**(1), 120–128 (2003).

¹⁸S. 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**(12), 2879–2890 (2007).

¹⁹D. 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* **54**(2), 177–184 (2006).

²⁰S. G. Armato, G. McLennan, M. F. McNitt-Gray, C. R. Meyer, D. Yankelevitz, D. R. Aberle, C. I. Henschke, E. A. Hoffman, E. A. Kazerooni, H. MacMahon, A. P. Reeves, B. Y. Croft, and L. P. Clarke, "Lung Image Database Consortium: Developing a resource for the medical imaging research community," *Radiology* **232**(3), 739–748 (2004).

²¹G. Gerig, M. Jomier, and M. Chakos, "Valmet: a new validation tool for assessing and improving 3D object segmentation," *Lect. Notes Comput. Sci.* **2208**, 516–523 (2001).

²²P. Aljabar, R. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert, "Classifier selection strategies for label fusion using large atlas databases," *Lect. Notes Comput. Sci.* **4791**, 523–531 (2007).

²³M. Prokop and M. Galanski, *Spiral and Multislice Computed Tomography of the Body* (Thieme, Stuttgart, 2003).