INTERACTIVE LUNG SEGMENTATION IN CT SCANS WITH SEVERE ABNORMALITIES

Thessa T.J.P. Kockelkorn[1]*, Eva M. van Rikxoort*[2]*, Jan C. Grutters*[3] *and Bram van Ginneken*[4,1]

[1]Image Sciences Institute, UMC Utrecht, The Netherlands
[2]Thoracic Imaging Research Group, University of California, Los Angeles, USA
[3]Department of Pulmonology, St Antonius Ziekenhuis Nieuwegein, The Netherlands
[4]Diagnostic Image Analysis Group, Department of Radiology,
Radboud University Nijmegen Medical Centre,The Netherlands

ABSTRACT

Estimation of the volume of the lungs and the viable lung tissue is an important step in the management of patients with severe pulmonary disease. The presence of gross pathology makes it impossible to perform lung segmentation automatically and reliably in CT scans of such patients. An interactive system for lung segmentation is presented, based on precomputed compact regions with homogeneous texture for which general texture feature have been computed. A statistical classifier trained on prior data has classified these regions beforehand and the user corrects any errors until the segmentation of an entire slice is correct. The system proceeds to subsequent slices, which were preclassified using a combined classification strategy that uses both the prior data and the previously approved slices from the test scan.

The resulting lung segmentations show a large overlap and a small average boundary distance when compared to completely manual delineations of the lung borders. The lung segmentations can then be used as input for a similar interactive system to determine the viable lung volume.

Index Terms— computed tomography, lung segmentation, interactive segmentation

1. INTRODUCTION

For many patients, clinicians are interested in an accurate estimation of the lung volume and the volume of viable lung tissue within the lungs. Computed tomography (CT) scans are the most accurate way to obtain this information. An important class are patients with acute lung injury/acute respiratory distress syndrome (ALI/ARDS), who require mechanical ventilation. Knowledge about the amount of viable lung tissue is essential to obtain appropriate ventilation settings, balancing the recruitment of collapsed lung tissue on the one hand and the prevention of ventilation induced lung injury on the other [1]. Gattinoni and Pesenti [2] describe the concept of a functional 'baby lung', the phenomenon that the normally aerated tissue in ALI/ARDS patients has the dimensions of the lungs of a 5- or 6-year-old child. They conclude that the smaller the baby lung, the greater the risk that mechanical ventilation is too aggressive and causes life threatening damage. However, reliable methods to estimate the baby lung volume are still to be found. Also for patients suffering from other pulmonary diseases, such as severe interstitial lung disease (ILD), cystic fibrosis or emphysema, and for patients that may be eligible for lung transplantation, measuring total and viable lung volume would be of great value in both determining the best way to treat the patient and in evaluating the treatment.

For this purpose, an accurate segmentation of the lungs in CT scans is a prerequisite. In the past, different segmentation methods have been proposed, which work well in relatively healthy lungs or even scans with pathology [3]. However, when large amounts of pathology are present, fully automatic lung segmentation is infeasible. In this work, an interactive system for segmenting the lungs and the viable lung tissue is presented. In the current study, we focus in particular on the evaluation of the lung segmentation.

2. MATERIALS

Twelve thoracic CT scans from lung transplantation and ILD patients were used. Scans were acquired at the University Medical Center Utrecht and St Antonius Ziekenhuis Nieuwegein, both in the Netherlands, between 2003 and 2006. Scans were made on either a Philips Mx8000 IDT scanner or a Brilliance 16P scanner (Philips Medical Systems, Best, The Netherlands) at full inspiration. No contrast material was used. Section thickness ranged from 0.9 to 1.0 mm and in plane resolution ranged from 0.6 to 0.9 mm. For training the interactive system, ten CT scans from a lung cancer screening trial were used. Scans were acquired with a similar protocol, only with a lower dose.

3. METHODS

The interactive annotation process is depicted schematically in Figure 1. In all scans, the body was segmented and divided

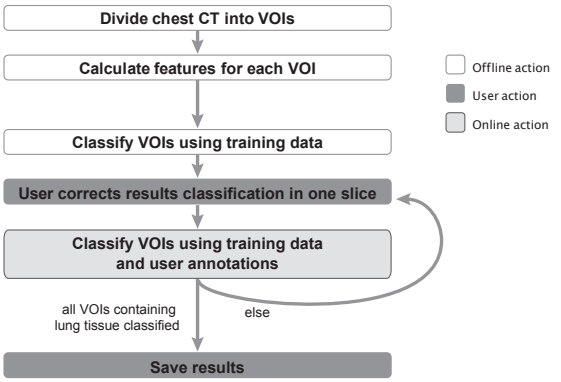


Fig. 1. Schematic overview of the interactive lung segmentation process

into compact volumes of interest (VOIs) with homogeneous texture. This was done by downsampling the images with a factor 2 in each dimension and subsequent blurring with a Gaussian kernel with σ = 1 voxel. Local gray value minima and maxima which were at least 8 voxels apart were used as seeds for growing the VOIs. To these seeds, all voxels within a radius of 3 voxels were added to form initial VOIs. For all voxels neighboring these initial VOIs, a dissimilarity score S was calculated, based on the distance from the neighboring voxel to the seed and the difference between the voxel density value and the average density of the initial VOI:

$$S = \alpha \left| (H_v - \overline{H}) \right| + d^2, \quad (1)$$

where H_v is the density of the voxel in Hounsfield units, $\overline{H}$ the average density value in the initial VOI and d the distance in voxels from the voxel to the center of the VOI. α denotes the relative weight assigned to the absolute difference in density. The voxel with the lowest dissimilarity score was added to its adjacent VOI and dissimilarity scores were computed for unassigned neighbors of the selected voxel. This process continued until all body voxels were assigned to a VOI. One scan contained between 10,000 and 19,000 VOIs. For each VOI in both the training and test images, texture features were calculated. Images were filtered using Gaussian, gradient magnitude and Laplacian filters at scales 1, 2, and 4 voxels. In these filtered images, the mean, standard deviation, skew and kurtosis of each VOI were calculated. This resulted in 36 texture features per VOI.

A human observer performed the interactive segmentation on each test scan. She was presented with one slice of the scan at a time, in which all VOIs intersecting with this slice had been labeled by the system as either lung tissue or non-lung tissue and she only had to correct wrongly labeled VOIs by mouse clicks. In this process, she could scroll through the scan to view the VOIs and their environment in all orthogonal directions. She was free at any time to label VOIs in other slices as well. The overlay with the labels of the VOIs could be switched on and off. The order in which the slices were presented was chosen in such a way to maximize the distance between already labeled slices.

To start the annotation process, all VOIs in the first slice were classified using the 160,000 training samples as input for a k-nearest neighbor (kNN) classifier (C_1, $k = 7$). The observer corrected the errors in this slice. By clicking a VOI with the left or right mouse button, the label of the VOI was set to respectively lung tissue or non-lung tissue. When done, the labels of all VOIs in this slice, both corrected and not corrected, were used to train a second kNN classifier (C_2, $k = 7$). When classifying the next slice, the algorithm compared the output of C_1 and C_2 for each VOI intersecting with the slice and, in case of disagreement, used the classifier that produced the highest posterior probability for lung or background. This combined classification strategy was used for all subsequent slices. The process continued until the observer decided that all VOIs had been labeled correctly.

All scans were also segmented using a region growing method described in [3]. To evaluate both segmentation methods, outlines of the lungs in several axial, coronal and sagittal slices in the test scans were drawn manually by a medical student. In each scan, in total 12-21 contours were drawn. These contours were compared to the results of both segmentation methods by calculating their overlap and their mean distance. The overlap O was defined as the number of voxels in the intersection of segmentation and reference divided by the number of voxels in their union. The mean distance MD was defined as the average of the distance for each contour pixel in one segmentation to the nearest contour pixel in the reference contour. Results were calculated per slice and per scan.

4. RESULTS

In Table 1, the results of the interactive segmentation method are shown. In six cases, shown in the first six rows of the table, the conventional segmentation method yielded unsatisfactory results, showing overlap values ranging from 0.57 to 0.89. In these cases, the interactive method performed better, yielding overlap values of 0.94-0.97 for all slices. The mean distance over all slices showed a similar pattern: 2.7-17mm for the conventional method versus 1.3-2.2 for the interactive segmentation.

In the middle and right part of Table 1, results are given per slice. The middle part shows the results for the slice that showed the lowest overlap between the manual delineation and the conventional lung segmentation. In these slices, the overlap for the conventional method ranges from 0.0 to 0.71 for scans in which the segmentation failed. The interactive method shows an overlap between 0.91 and 0.99 for the same slices. The mean distance values show the same trend: for the interactive method, they range between 0.7 and 3.0mm, for the conventional method between 5.0 and 33mm. In the right

Table 1. Results of interactive lung segmentation and lung segmentation by region growing compared to manual lung border delineation. The upper rows (Improvement) list the results for scans where the conventional lung segmentation did not produce satisfactory results. The lower rows (No improvement) list results for scans for where both methods yielded good results. In the first four columns, results are given for the entire scan. The middle four columns give the results for the slice that performed worst in the conventional segmentation and the last four columns give the results for the slice that performed worst in the interactive segmentation. For all these three blocks, results are given for the interactive segmentation method (interactive) and the region growing method (conventional). Outcome measures are the overlap (O) and the mean distance MD, in mm. N/A indicates the mean distance could not be calculated.

	average over all slices				worst slice conventional				worst slice interactive			
	interactive		conventional		interactive		conventional		interactive		conventional	
	O	MD	O	MD	O	MD	O	MD	O	MD	O	MD
Improvement	0.94	1.3	0.85	2.9	0.96	0.8	0.72	5.3	0.87	3.4	0.73	5.6
	0.95	1.9	0.72	7.1	0.92	2.2	0.11	32	0.90	4.3	0.82	5.3
	0.96	1.7	0.62	12	0.91	3.0	0.00	N/A	0.91	3.0	0.00	N/A
	0.97	1.8	0.57	17	0.99	0.7	0.12	31	0.92	3.1	0.60	9.9
	0.96	2.2	0.76	6.8	0.98	1.0	0.40	33	0.87	5.3	0.89	4.2
	0.96	1.5	0.89	2.7	0.91	2.5	0.69	5.0	0.91	2.5	0.69	5.0
No improvement	0.96	1.7	0.97	1.6	0.92	2.9	0.94	2.2	0.92	1.8	0.95	1.2
	0.96	1.4	0.97	1.0	0.91	2.5	0.93	1.8	0.91	2.5	0.93	1.8
	0.97	1.4	0.98	1.0	0.95	6.7	0.96	4.8	0.95	6.7	0.96	4.8
	0.96	1.8	0.97	1.2	0.91	5.4	0.94	4.2	0.91	5.4	0.94	4.2
	0.96	1.8	0.97	1.2	0.92	2.4	0.92	2.1	0.92	2.4	0.92	2.1
	0.96	1.7	0.97	1.3	0.94	7.7	0.94	6.6	0.94	7.7	0.94	6.6

part of table 1, results are shown for the slices that showed the lowest overlap in the interactive method. In one of these cases, the conventional method outperformed the interactive method in segmenting this slice, with overlap values of 0.89 versus 0.87 and mean distances of 4.2 and 5.3 mm respectively. In all the other cases, the interactive method outperformed the conventional method.

Figure 2 shows the different segmentations for an axial slice in which the conventional method did not perform well. This slice shows dense areas in the dorsal part of both lungs (a). The manual segmentation is shown in panel (b). The conventional lung segmentation was not able to deal with these abnormalities and left them out of the segmentation. The interactive method captured the entire lung field in this slice and showed a high correspondence to the manual delineation.

In the cases where the conventional segmentation was already good, as shown in the lower part of table 1, the interactive method also yielded good results, but the performance of the conventional method is consistently better. This difference might be due to the fact that the interactive method uses predefined VOIs. If a VOI contains both lung tissue and non-lung tissue, part of its voxels are misclassified. This may cause a less accurate segmentation. The use of smaller VOIs may increase the accuracy of the interactive method, but this would probably increase the user interaction effort.

Results for the segmentation time and number of corrections needed to yield a complete segmentation are given as average (standard deviation) over all scans. The time needed for a complete segmentation was 10 (3.3) minutes. As a comparison, a manual delineation of one lung in one slice takes 1.2 minutes [3]. The human observer needed to adjust the label of 168 (71) VOIs per scan. This corresponds to 1% of all VOIs in a scan. The number of adjustments needed to segment a scan in which the automatic segmentation failed was higher than the number of corrections needed in a scan in which conventional segmentation was successful; 175 (87) versus 160 (58). This can be explained by the amount of dense abnormalities in the scans: if a larger area of the lungs contains dense textures, the conventional lung segmentation is more likely to fail there. Since the interactive segmentation method is initially trained with data from different subjects, it is less likely to recognize areas with gross pathology as lung tissue, and the user needs to make more corrections. These numbers indicate that the interactive annotation process is a user-friendly way to yield a reliable lung segmentation.

Figure 3 shows a screenshot of the application that uses the output of the lung segmentation to estimate the viable lung tissue volume. The system for interactively annotating different textures in thoracic CT scans has been described in [4]. It uses all VOIs that have been labeled as lung tissue. The user starts with annotating several VOIs of each texture present, in this case viable and non-viable lung tissue. Based on these labels and the texture features described before, a kNN classifier classifies all VOIs in the next slice. The user corrects the mistakes and the system uses the new information to classify the next slice. This process continues until all VOIs have been classified. The volume of lung tissue of each class, in this case viable and non-viable lung tissue, is then calculated.

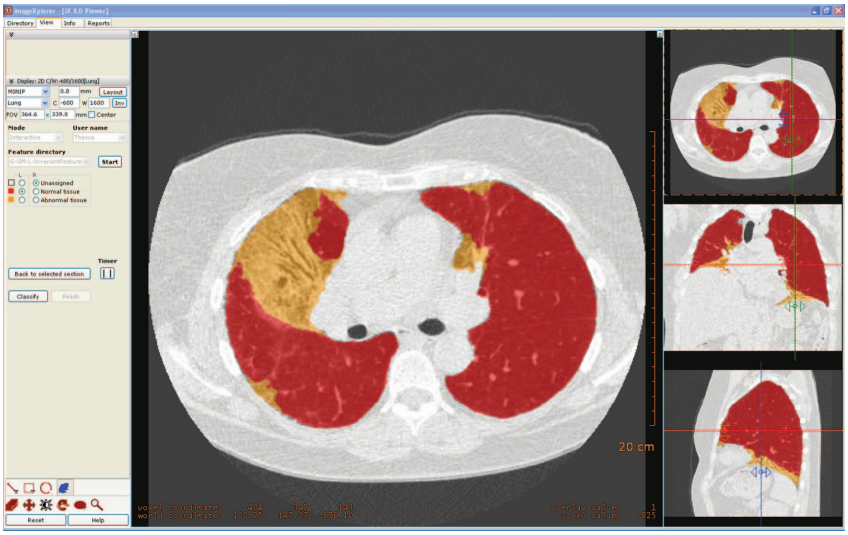
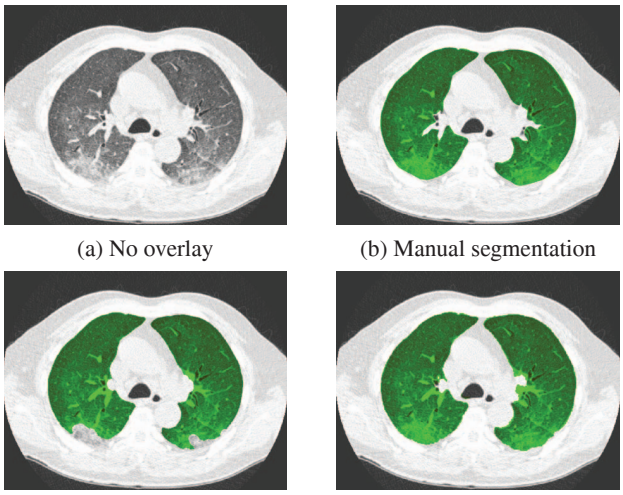


Fig. 3. Example of classification of lung VOIs as normal (red) or abnormal (light brown) lung tissue



(a) No overlay (b) Manual segmentation

(c) Conventional segmentation (d) Interactive segmentation

Fig. 2. Example of the results for a coronal slice (a) in which the conventional lung segmentation algorithm failed (c). Parts of the lung were missed or misclassified as airways. The manual delineation was converted to an overlay and shown in (b). Results of the interactive segmentation method are shown in (d). They show a good correspondence to the manual segmentation.

5. CONCLUSION AND DISCUSSION

The proposed method can be used to segment lungs in scans in which automatic segmentation methods fail. Future work includes the optimization of the system so that classification of VOIs using C_2 is performed in separate processing threads and is instantaneous. Also other classifier combination strategies will be investigated. It may be beneficial to extend the feature set with spatial features and shape features. It could also be advantageous to include the results of conventional lung segmentation in the initial classification from C_1. Finally, we intend to use a similar strategy with two classifiers C_1 and C_2 for the second step of determining the viable lung tissue among the segmented lung VOIs.

We expect that these improvements will bring down the required interaction time to no more than a few minutes, allowing application of the system in clinical routine. And although we do not expect that for the accurate estimation of lung volume it is necessary, it might be useful to allow the user to split any VOIs that do not follow the lung borders exactly.

6. REFERENCES

[1] P. Caironi, T. Langer, and L. Gattinoni, "Acute lung injury/acute respiratory distress syndrome pathophysiology: what we have learned from computed tomography scanning," *Current Opinion in Critical Care*, vol. 14, no. 1, pp. 64–69, 2008.

[2] L. Gattinoni and A. Pesenti, "The concept of "baby lung"," *Intensive Care Medine*, vol. 31, no. 6, pp. 776–784, 2005.

[3] E. M. van Rikxoort, B. de Hoop, M. A. Viergever, M. Prokop, and B. van Ginneken, "Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection," *Medical Physics*, vol. 36, no. 7, pp. 2934–2947, 2009.

[4] T. T. J. P. Kockelkorn, P. A. de Jong, H. A. Gietema, J. C. Grutters, M. Prokop, and B. van Ginneken, "Interactive annotation of textures in thoracic CT scans," in *SPIE Medical Imaging, Accepted*, 2010.