Lung Segmentation from CT with Severe Pathologies Using Anatomical Constraints

Neil Birkbeck[1], Timo Kohlberger[1], Jingdan Zhang[1],
Michal Sofka[1], Jens Kaftan[2], Dorin Comaniciu[1], and S. Kevin Zhou[1,⋆]

[1] Imaging & Computer Vision, Siemens Corporate Technology, Princeton, NJ, USA
[2] Molecular Imaging, Siemens Healthcare, Oxford, UK

Abstract. The diversity in appearance of diseased lung tissue makes automatic segmentation of lungs from CT with severe pathologies challenging. To overcome this challenge, we rely on contextual constraints from neighboring anatomies to detect and segment lung tissue across a variety of pathologies. We propose an algorithm that combines statistical learning with these anatomical constraints to seek a segmentation of the lung consistent with adjacent structures, such as the heart, liver, spleen, and ribs. We demonstrate that our algorithm reduces the number of failed detections and increases the accuracy of the segmentation on unseen test cases with severe pathologies.

1 Introduction

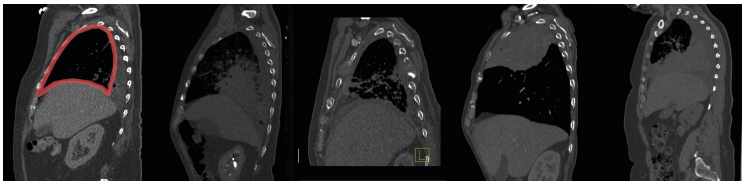
Healthy lung tissue is easily distinguishable from surrounding soft tissue in CT images due to its high air content. In contrast, pathologies like tumors, interstitial lung diseases (ILD), and plural effusion [1], may dramatically alter both the appearance and texture of the lung and its surroundings (see Fig. 1). Due to this variability, automatic methods for segmenting healthy lung that rely on appearance alone, such as region growing or registration, are inappropriate for segmenting lungs from CT with severe pathologies [7].

In order to account for specific pathologies, such as ILD, some existing methods focus on utilizing texture cues [4,10], or for tumors, robust statistical shape models can be used [9]. In order to robustly segment diseased lung parenchyma, other researchers have identified the need to use other nearby anatomical context. For example, the curvature of nearby ribs [6] or the distance to spine [2] can be used in addition to appearance information, or stable landmarks that take into account neighboring ribs can be used to better align a statistical shape model on the lung [8].

⋆ Zhou is corresponding author. Birkbeck and Kohlberger are with Google, Zhang with Microsoft, and Sofka with Cisco. All work was done while they were with Siemens.

[1] Pleural effusion is an accumulation of pleural fluid between the parietal and visceral pleura, and thus part of the pleural cavity, but not of the actual lung. It is a common cause for decreased lung function and only affects the lungs directly. For this reason, we include both the lung and the pleural cavity in our segmentation.

P. Golland et al. (Eds.): MICCAI 2014, Part I, LNCS 8673, pp. 804–811, 2014.
© Springer International Publishing Switzerland 2014



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

Fig. 1. Examples of pathological lungs. In (a), the boundary between low/high intensity gives rise to a lung-like shape (red), making it hard to rely solely on a shape prior to segment pathologies.

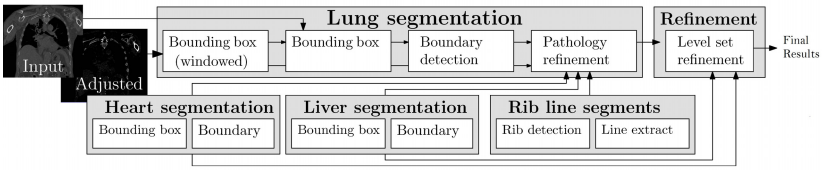


Fig. 2. The proposed pathological lung segmentation pipeline

However, none of these methods utilize all available anatomical context surrounding the lung regions. We propose a learning-based algorithm capable of segmenting lung from CT scans with several pathologies. To be robust against pathologies, we rely on context external to the lung for each phase of the pipeline, including detection, coarse segmentation, and fine-scale refinement. For the segmentation, we use region-specific anatomical constraints such as distance to ribs or adjacent organs (e.g. liver and heart).

2 Methods

The lung is bounded by the ribs on the outside surface, the heart on the mediastinum side, and, e.g., the right lung is bounded by the liver on the bottom. In order to use these adjacent structures, our algorithm starts by automatically detecting them (e.g., lung, heart, liver & spleen §2.1) and rib segments (§2.2) using statistical classifiers. The lung mesh is then deformed using appearance cues and neighboring structures as geometric constraints (§2.3). A final refinement balances the need for a fine-scale geometry while maintaining proximity to adjacent structures and ensuring no overlap between segmentations (§2.5). This final refinement maintains the integrity of the segmentation in the presence of pathologies by constraining the refined solution to lie close to the detected solution. The algorithm components and data flow are outlined in Fig. 2.

2.1 Organ Detection

A central assertion in our approach is that we can apply region-specific geometric constraints to the segmentation. In order to apply such constraints, we need our

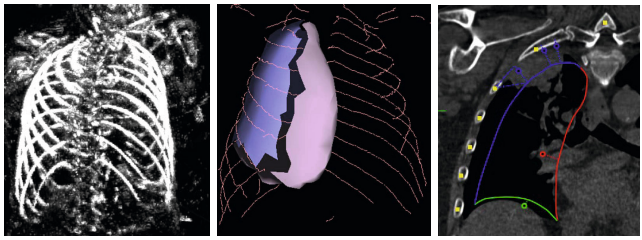


Fig. 3. Examples of rib lines (left), the 3D region parts for the right lung (middle), and region specific geometric constraints (right). The lung surface is divided into three partitions: the rib region (blue), the heart/mediastinum region (red) and the bottom region (green). Each region of the surface uses neighboring anatomy as constraints.

detected surfaces to be in correspondence with a template surface. To this end, we initialize our organ segmentation through the use of a statistical shape model that has been learned from training data.

Let $\mathcal{C} = (\mathcal{V}, \mathcal{T})$ be a mesh with vertices $\mathcal{V} = \{\mathbf{v}_i = (x_i, y_i, z_i)\}_{i=1}^{n}$ and triangles $\mathcal{T}$. We model shape variation with a point-distribution model [1] using a low-dimensional set of M shape modes, $\mathbf{U}_j = \{\mathbf{u}_{ij}\}_{i=1}^{n}$, and a mean vector $\bar{\mathbf{v}}_i$: $\mathbf{v}_i = \bar{\mathbf{v}}_i + \sum_{j=1}^{M} \lambda_j \mathbf{u}_{ij}$, where $\{\lambda_j\}_{j=1}^{M}$ is a shape coefficient vector defining the shape. Given a transformation, $T(\mathbf{v}; \theta)$ parameterized by θ, e.g., a similarity transform, the initial organ detection seeks the most probable configuration given the input image, I:

$$\mathcal{C}^0 = \text{argmax}_{\mathcal{C}}\ P(\mathcal{C}|I, \{\bar{\mathbf{v}}_i\}, \{\mathbf{u}_{ij}\}) = \text{argmax}_{\mathcal{C}}\ P(\theta, \{\lambda_j\}_{j=1}^{M}|I, \{\bar{\mathbf{v}}_i\}, \{\mathbf{u}_{ij}\}). \quad (1)$$

This problem is solved with the discriminative method marginal space learning [11], and it is used to reliably estimate an initial object segmentation of the lungs, $\mathcal{C}^0_{\text{llung}}\&\mathcal{C}^0_{\text{rlung}}$, heart, $\mathcal{C}^0_{\text{heart}}$, spleen, $\mathcal{C}^0_{\text{spleen}}$, and liver, $\mathcal{C}^0_{\text{liver}}$, (e.g., [5]).

For the lungs, to avoid failed detection on large effusions with dramatically different appearance, we first estimate the shape on a contrast adjusted input image, $I_{win} = \max(0, \min(1, (I - 80)/496))$, where I is in Hounsfield unit (HU).

2.2 Rib Points Extraction

In order to extract points on the ribs to be used as constraints, discriminative classifiers are used to identify likely rib points. These likely rib points are then grouped into short segments, and unlikely segments are culled (Figure 3).

First, a per-voxel rib probability map, $I_{rib}(\mathbf{x})$, is obtained from a two-level cascade of discriminative classifiers,

$$I_{rib}(\mathbf{x}) = P^{\text{rib}}_{2\text{mm}}(+1|I_{2mm}, \mathbf{x})\pi[P^{\text{rib}}_{2\text{mm}}(+1|I_{2mm}, \mathbf{x}) > \tau_2]\pi[P^{\text{rib}}_{4\text{mm}}(+1|I_{4mm}, \mathbf{x}) > \tau_1] \quad (2)$$

where $\pi[.]$ is an indication function, P^{rib} are rib point classifiers trained using Haar-like features, τ_1 is a threshold used to limit computation in the 2nd level

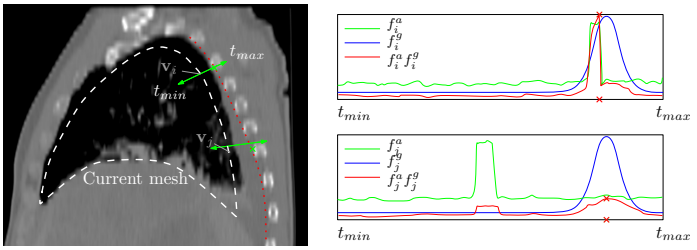


Fig. 4. Two points on the current mesh surface. $\mathbf{v}_i$ is a healthy point, so the appearance term is reliable, and the geometric term has little effect as indicated by the response curves as a function of depth. On the other hand, $\mathbf{v}_j$ is a pathological point and the appearance term is misleading. The product of appearance and geometry gives the correct displacement.

(2mm) to highly likely voxels, and τ_2 [2] is used to further threshold the 2mm classifier response. I_{4mm} and I_{2mm} are volumes that have been isotropically resampled to the corresponding resolution.

From the rib probability map I_{rib}, connected components of the thresholded mask are obtained. Each segment is grown to a maximum of size of $s_{max} = 9000\text{mm}^3$. Components less than $s_{min} = 2000\text{mm}^3$ are pruned. The remaining segments should be like bent cylinders. To prune candidate segments that are not like bent cylinders, a final test on the sizes of the segment along its principle directions is performed. Let $\alpha_1 > \alpha_2 > \alpha_3$ denote the sizes of the segment in its three principle directions. Remaining segments are pruned if $\frac{\alpha_1}{\alpha_2} < 2.5$ or $\frac{\alpha_2}{\alpha_3} > 3$.

Finally, the center points of each segment are extracted by traversing along its principal direction and intersecting a fixed number of planes with the segment.

2.3 Pathological Lung Surface Refinement

The key component of the pathological boundary detector is to combine image appearance cues with spatially dependent anatomical constraints that come from the ribs and organs detected using methods in the previous sections. The mesh initialization from §2.1 is already in correspondence with a known atlas. Also, the indices, $\mathcal{I} = \{1 \leq i \leq n\}$ of the mean mesh vertices are partitioned into 3 subsets: $\mathcal{I} = \mathcal{I}_{ribs} \bigcup \mathcal{I}_{heart} \bigcup \mathcal{I}_{bottom}$, each subset corresponding to one anatomical region. During boundary refinement, each vertex independently is deformed to $\mathbf{v}_i \leftarrow \mathbf{v}_i + t_i\mathbf{n}$, where

$$t_i = \text{argmax}_{t \in [t_{\min}, t_{\max}]} P_i(t|I), \tag{3}$$

with the search being locally constrained within a range $[t_{\min}, t_{\max}]$. The resulting displaced mesh is regularized by projection onto the shape subspace, and the process is repeated for a few iterations.

The per-vertex score in (3) is computed as a combination of appearance, $f_i^a(t)$, and structural cues, $f_i^g(t)$, $P_i(t|I) = f_i^a(t) f_i^g(t)$. We assume $f_i^a(t)$ is the

[2] In practice, we tune τ_1 and τ_2 through the ROC curve and the final performance is insensitive to small changes in these parameters.

best model to use assuming no pathology. However, the opposite is true for pathological surfaces, where appearance cues are misleading, and we rely on $f_i^g(t)$. Figure 4 illustrates how the geometric term helps in the case of a pathology.

Unlike previous methods that model the appearance term with an image gradient (e.g., [9]), or region-based term, we instead use a discriminative model for the appearance in the form of a classifier that is learned from training data [5]:

$$f_i^a(t) = P(+1|\mathbf{v}_i + t\mathbf{n}_i, I). \tag{4}$$

The classifier, $P(+1|\mathbf{v}_i + t\mathbf{n}_i, I)$, measures the probability that the displaced point $\mathbf{v}_i + t\mathbf{n}_i$ belongs to the surface of the lung. To model this posterior, we use a probabilistic boosting tree with steerable features [11].

The anatomical term, $f_i^g(t)$, penalizes the surface from being far away from the adjacent structure using a Gaussian kernel mixed with a uniform prior:

$$f_i^g(t) = \exp\left(-\frac{(t - (\mathbf{y}_{\text{struct}}(\mathbf{v}_i) - \mathbf{v}_i)^T\mathbf{n}_i - \tau_i)^2}{2\sigma_i^2}\right) + \frac{\alpha}{t_{max} - t_{min}}, \tag{5}$$

where $\mathbf{y}_{\text{struct}}(\mathbf{v})$ is the closest point on the adjacent relevant structure from the source point $\mathbf{v}$. The per-vertex τ_i parameter allows for a spatially varying offset between the adjacent structure and the lung surface being segmented. These spatially varying parameters account for errors in mesh correspondence and unreliability of geometric constraints, e.g., on regions close to the spine. Adding the uniform prior reduces the influence of the geometric in the case that $\mathbf{y}_{\text{struct}}(\mathbf{v})$ is far away from $\mathbf{v}_i$.

We estimate the spatial varying $\{\tau_i\}$ and $\{\sigma_i\}$ from a given set of K training instances. We first compute for each vertex i an estimate of τ_i as the average distance to the neighboring structure from that point: $\bar{\tau}_i = \frac{1}{K}\sum_k^K \tau_{i,k}$, with $\tau_{i,k}$ being the distance estimated for training sample k. The standard deviation can be used as an estimate for σ_i. To ensure these values vary smoothly over the surface, we perform a weighted smoothing operation over the surface of the mesh, $\tau_i = (\bar{\tau}_i + mean_{j\in N(i)}(\bar{\tau}_j))/2$, where $N(i)$ are neighbors of vertex i.

2.4 Identifying Pathological Lungs

To make our pipeline capable of also efficiently handling healthy lung cases, we only apply the pathological boundary processing if the initialization from the basic pipeline is determined to be pathological. Determining if a case is healthy or pathological also uses anatomical constraints. A lung is determined as pathological when either the anatomical constraints between the lung/rib surface and lung/bottom organ surface are not satisfied. (Fig. 5).

For the lung/rib surface, first, we sample the rib probability image above the surface of the lung attached to the rib and take the maximum value of I_{rib} at values of 2mm, 4mm, and 6mm above the lung surface. We then produce a binary function on the lung surface for regions of the lung surface that are confidently close to a rib as $\chi(\mathbf{v}) = (\max_{t=2,4,6} I_{rib}(\mathbf{v} + \mathbf{n}t)) > 0.67$. If any part of the rib-adjacent lung surface is more than 25mm from a point with $\chi(\mathbf{v}) = 1$,

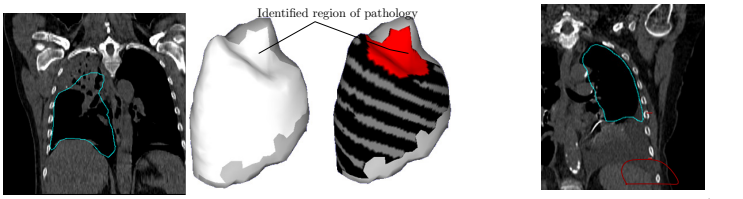


Fig. 5. If lung regions adjacent to the rib surface are too far from the ribs, the input is identied as pathological. For the bottom region, the pathological pipeline will be activated if the surface is not close enough to the adjacent organ (e.g., liver or spleen).

then we declare that the lung is a pathology. Figure 5 illustrates the regions of the surface near a rib, as well as the red region farther than 25mm.

To identify cases where the bottom region is pathological, the vertices on the bottom region that are closely coupled to the bottom organ are identified as $A = \{\mathbf{v}_i : i \in \mathcal{I}_{bottom}, \sigma_i \leq 5\}$. If too many vertices in A (e.g., 15%) are far from the bottom organ (say beyond $3\sigma_i$), we perform the pathological processing.

2.5 Final Refinement

Although the explicit mesh-based representation above allowed us to easily define region-specic anatomical constraints that were necessary to segment pathological lungs, the mesh-based surface representation is inadequate in obtaining voxel-level fine-scale detail. In order to obtain a more precise segmentation, we convert the initialized mesh surfaces into an implicit formulation and perform a fine-scale refinement using level sets [3]. The level set representation allows us to integrate appearance-based data terms with anatomical constraints (such as non-overlapping constraints) between neighboring organs. See [3] for details.

3 Experiments

Our lung detectors and related parameters have been trained on a total of 185 right lung and 196 left lung cases. Using the percentage of ground truth lung volume with values greater than 512 HU as an indicator of the severity of the pathology, the majority of our training examples are from healthy lungs: only 7 (resp. 2) left (resp. right) lung cases had between 25%-50% unhealthy tissue. We test our pipeline on a set of CT images containing 17 left and 33 right lung pathologies having an average dimension of $465 \times 448 \times 328$ and spacing of $1.11 \times 1.11 \times 2.61$mm. Our cases have 12 left lungs with greater than 25% unhealthy tissue, 5 of which are greater than 50%. There are 21 right lung cases with greater than 25% unhealthy tissue and 10 cases with greater than 50% unhealthy tissue. The average run time for our test cases is 64s, including processing time for lung, liver, heart, and spleen, and ribs. All training and test data sets are manually annotated by experts.

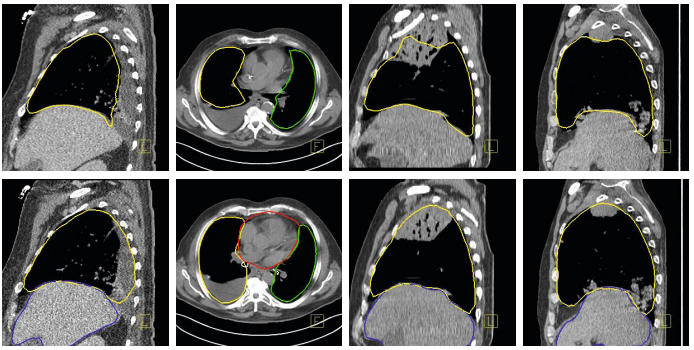


Fig. 6. Lung segmentation without (top) and with anatomical constraints (bottom) using the heart (red) and liver (blue). Errors decrease from 8.55mm to 4.49mm, 22.26mm to 6.16mm, 4.7mm to 3.3mm, and 3.2mm to 2.57mm.

Table 1. Average surface-to-surface error (Mean, median, and 80% highest error in mm) & volumetric stats (Dice Coeff and Relative Volume Difference) on unseen cases.

	Mean(STD)	Median	80%	Dice Coeff	Rel.Vol.	Fail
L.Lung Basic	4.76(2.47)	4.23	7.18	0.84	-0.14	3
L.Lung Path.	3.44(0.96)	3.84	4.25	0.89	-0.06	0
L.Lung LevSet	2.49(1.14)	2.19	3.69	0.91	0.01	0
R.Lung Basic	5.72(3.96)	4.02	8.25	0.85	-0.19	5
R.Lung Path.	3.64(1.21)	3.36	4.31	0.91	-0.06	0
R.Lung LevSet	2.23(0.97)	2.01	2.69	0.94	0.06	0

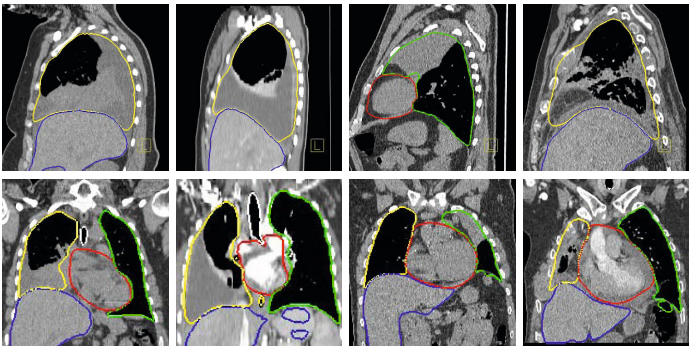


Fig. 7. Qualitative results of the hard cases used in the experimental analysis

The basic detection pipeline that relies on the input image only to detect the bounding box fails on 3/17 and 5/33 cases for left and right lung respectively (i.e., bounding box fails to return a box with high enough confidence). Table 1 illustrates statistics on mean surface-to-surface distance error (in mm) and volumetric statistics for the unseen test set. The error was only computed on the non-failure cases. In the table, the mean and median errors decrease after using the pathological pipeline. But as pathologies may only occupy a small region of

the surface, the error measure between the basic detection and pathological detection may only decrease slightly. In the case of large tumors and effusions, the use of the anatomical constraints is essential to providing a geometric consistent lung segmentation. This is better illustrated with examples (Fig. 6). The geometric constraints give a segmentation that hugs the rib boundary regardless of appearance. The final refinement causes a decrease in surface error on all regions (including healthy regions), further boosting the performance. Fig. 7 shows the final level set refinement on several challenging cases.

4 Conclusions

The lung image appearance is affected by large changes due to pathologies. We have shown that structural constraints are necessary to accurately segment pathological lungs, when combined with statistical learning. Future improvements include more accurate rib segmentation, better correspondence in shape models, or other anatomical constraints say from the aorta.

References

1. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape models-their training and application. Computer Vision and Image Understanding 61(1), 38–59 (1995)
2. Hua, P., Song, Q., Sonka, M., Hoffman, E., Reinhardt, J.: Segmentation of pathological and diseased lung tissue in ct images using a graph-search algorithm. ISBI (March 2011)
3. Kohlberger, T., Sofka, M., Zhang, J., Birkbeck, N., Wetzl, J., Kaftan, J., Declerck, J., Zhou, S.K.: Automatic multi-organ segmentation using learning-based segmentation and level set optimization. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 338–345. Springer, Heidelberg (2011)
4. Korfiatis, P., Kalogeropoulou, C., Karahaliou, A., Kazantzi, A., Skiadopoulos, S., Costaridou, L.: Texture classification-based segmentation of lung affected by interstitial pneumonia in high-resolution CT. Medical Physics 35, 5290 (2008)
5. Ling, H., Zhou, S.K., Zheng, Y., Georgescu, B., Suehling, M., Comaniciu, D.: Hierarchical, learning-based automatic liver segmentation. CVPR, 1–8 (2008)
6. Prasad, M., Brown, M., Ahmad, S., Abtin, F., Allen, J., da Costa, I., Kim, H., McNitt-Gray, M., Goldin, J.: Automatic segmentation of lung parenchyma in the presence of diseases based on curvature of ribs. Acad. Radiol. 15, 1173–1180 (2008)
7. Sluimer, I., Prokop, M., van Ginneken, B.: Toward automated segmentation of the pathological lung in ct. TMI 24(8), 1025–1038 (2005)
8. Sofka, M., Wetzl, J., Birkbeck, N., Zhang, J., Kohlberger, T., Kaftan, J., Declerck, J., Zhou, S.K.: Multi-stage learning for robust lung segmentation in challenging CT volumes. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 667–674. Springer, Heidelberg (2011)
9. Sun, S., McLennan, G., Hoffman, E.A., Beichel, R.: Model-based segmentation of pathological lungs in volumetric ct data. In: The Third International Workshop on Pulmonary Image Analysis (2010)
10. Wang, J., Li, F., Li, Q.: Automated segmentation of lungs with severe interstitial lung disease in CT. Medical Physics 36, 4592 (2009)
11. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Fast automatic heart chamber segmentation from 3D CT data using marginal space learning and steerable features. In: ICCV. pp. 1–8. IEEE (2007)