# Multi-stage Learning for Robust Lung Segmentation in Challenging CT Volumes

Michal Sofka[1], Jens Wetzl[1], Neil Birkbeck[1], Jingdan Zhang[1], Timo Kohlberger[1], Jens Kaftan[2], Jérôme Declerck[2], and S. Kevin Zhou[1]

[1] Image Analytics and Informatics, Siemens Corporate Research, Princeton, NJ, USA
[2] Molecular Imaging, Siemens Healthcare, Oxford, UK

**Abstract.** Simple algorithms for segmenting healthy lung parenchyma in CT are unable to deal with high density tissue common in pulmonary diseases. To overcome this problem, we propose a multi-stage learning-based approach that combines anatomical information to predict an initialization of a statistical shape model of the lungs. The initialization first detects the carina of the trachea, and uses this to detect a set of automatically selected stable landmarks on regions near the lung (e.g., ribs, spine). These landmarks are used to align the shape model, which is then refined through boundary detection to obtain fine-grained segmentation. Robustness is obtained through hierarchical use of discriminative classifiers that are trained on a range of manually annotated data of diseased and healthy lungs. We demonstrate fast detection (35s per volume on average) and segmentation of 2 mm accuracy on challenging data.

## 1 Introduction

Lung segmentation in thoracic CT images is an important prerequisite for detection and study of the progression and treatment of pulmonary diseases. Due to their high air content, healthy lung has lower attenuation than the surrounding tissue, allowing easy detection through standard thresholding and region-growing methods (e.g., [2]). However, pulmonary diseases (e.g., pulmonary fibrosis) lead to higher density tissue, and cause a changed appearance (e.g., different texture), making it hard to segment robustly (Figure 1).

In this paper, we present an effective learning-based segmentation technique that addresses the changes in lung appearance due to pulmonary diseases. The first step of the algorithm is the robust detection of the carina of the trachea with a discriminative classifier. The carina location is used to *predict* approximate poses (translation, orientation, and size) of the left and right lung. The prediction is based on a prior model obtained from a large expert-annotated database of

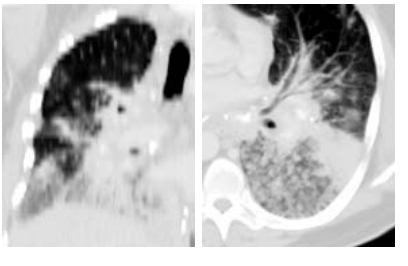

**Fig. 1.** Pulmonary diseases lead to higher density tissue which complicates standard segmentation algorithms

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part III, LNCS 6893, pp. 667–674, 2011.

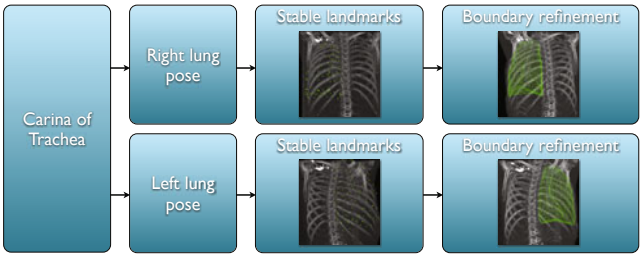


**Fig. 2.** System diagram: Carina detection allows the prediction of lung poses, which give initial locations for stable landmarks on the lung surface. That surface is then refined with a boundary detector.

lung scans. Placing a mean lung shape into the bounding box implied by each pose gives initial locations of a set of stable landmarks, which are selected automatically during training using the uncertainty of their locations. The locations of the landmarks are then locally refined with robust detectors. The final refinement is performed by a boundary detector which accurately estimates the lung surface. The overall system diagram is shown in Figure 2.

By focusing on stable landmarks and progressing from a coarse set to to a fine set, we rely on the local regions which are most consistent, even in the presence of abnormalities. Stable landmarks are typically selected near vertebrae, ribs, and other distinctive anatomical structures (Fig. 4). In detection, the stability of the landmarks is further guaranteed by using a discriminative classifier (PBT [10]) which includes a powerful feature selection operating on a feature pool computed from all training volumes. Features susceptible to abnormal diseased areas are therefore not selected for landmark detection.

Existing approaches to increase robustness of lung segmentation focus on specific pathologies [11,4], rely on interaction [3], adapt the simple thresholding to regions that often complicate lung segmentation [1], or augment texture cues for interstitial lung disease [12]. Such methods are not capable of handling the moderate to extreme pathologies that exhibit higher density tissue. More elaborate methods use anatomical information [6], shape priors [9], or statistical methods to detect patterns in the diseased tissue [7]. Shape models alone with simple image cues [9] are not enough to provide robustness to change in tissue density, and the anatomical constraints, priors or machine learning techniques need to be combined. To date, few methods combine either shape constraints or anatomical information with learning [7] for lung segmentation, but do so with simple classifiers (e.g., k-nearest neighbor) and use limited features.

In this work, we combine a statistical model of shape variation with statistical pattern recognition that uses anatomical information for robust lung segmentation. A wide gamut of gradient and intensity features capable of discriminating diseased lung tissue and implicitly capable of encoding anatomical relationships is selected by a powerful discriminative classifier, the probabilistic boosting tree (PBT) [10]. The classifiers and shape model are trained on a database of normal

and diseased tissues. Through fast coarse detection and refinement based on a hierarchical detection network (HDN) [8], a segmentation is obtained in 35 seconds. We demonstrate our implementation on a number of challenging pathological thoracic CT images. The average error on unseen data is 1.98 mm for the right and 1.92 mm for the left lung.

## 2 Learning

The algorithm starts by detecting the carina of the trachea. Its location is then used in a Hierarchical Detection Network (HDN)[8] to predict pose parameters of left and right lung and subsequently initialize the set of stable landmarks. The landmark locations are then refined and used to provide a rough boundary estimate. This estimate is improved during boundary detection which results in the final accurate lung segmentation.

### 2.1 Hierarchical Detection Network

The hierarchical detection network estimates unknown object states (e.g., object poses) as a sequential decision process. The formulation is similar to Markov chain approaches to object tracking, but instead of a temporal motion model with temporal observations, there is a spatial dependence (or prior relationship) between objects. The unknown parameters of each object are denoted as $\boldsymbol{\theta}_t$ (e.g., the 9 parameters of a similarity transform), and the complete state for $t+1$ objects is denoted $\boldsymbol{\theta}_{0:t}$. Estimation of each object utilizes an observation region of the input volume, $V_t \subseteq V$, where $V : \mathbb{R}^d \mapsto \mathbb{R}$ is $d$ dimensional input. The posterior density of the complete state, $f(\boldsymbol{\theta}_{0:t}|V_{0:t})$, is approximated through a sequence of recursive *prediction* and *update* steps.

The *prediction* approximates the detection up to object $t$ using the transition probability, $f(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{0:t-1})$, and the posterior up to object $t-1$:

$$f(\boldsymbol{\theta}_{0:t}|V_{0:t-1}) = f(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{0:t-1})f(\boldsymbol{\theta}_{0:t-1}|V_{0:t-1}) \tag{1}$$

The *update* then fuses the results with the new observation region, $V_t$:

$$f(\boldsymbol{\theta}_{0:t}|V_{0:t}) = \frac{f(V_t|\boldsymbol{\theta}_t)f(\boldsymbol{\theta}_{0:t}|V_{0:t-1})}{f(V_t|V_{0:t-1})} \tag{2}$$

The likelihood, $f(V_t|\boldsymbol{\theta}_t)$, is empirically modeled by training a discriminative model. Concretely, letting $y \in \{-1, 1\}$ be a random variable denoting the occurrence of an object at pose $\boldsymbol{\theta}_t$, the likelihood is defined as:

$$f(V_t|\boldsymbol{\theta}_t) = f(y = +1|V_t, \boldsymbol{\theta}_t) \tag{3}$$

Where the posterior, $f(y = +1|V_t, \boldsymbol{\theta}_t)$, is the output of a discriminative classifier (e.g., the probabilistic boosting tree [10]).

The transition prior approximates the sequential dependence of object $t$ as a Gaussian distribution from one of the previous objects (Figure 2):

$$f(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{0:t-1}) = f(\boldsymbol{\theta}_t|\boldsymbol{\theta}_j), \quad \exists j \in \{0, t-1\} \tag{4}$$

### 2.2 Pose Detection

In the case of pose detection in 3D, the state of each object is compactly represented with 9 parameters including the position $\mathbf{p} \in \mathbb{R}^3$, orientation as Euler angles $\mathbf{r}$, and scale, $\mathbf{s}$, of the object: $\boldsymbol{\theta}_t = \{\mathbf{p}_t, \mathbf{r}_t, \mathbf{s}_t\}$. For efficiency these three sets of parameters are treated as a chain of dependent estimates [5]:

$$f(\boldsymbol{\theta}_t|V_t) = f(\mathbf{p}_t|V_t) f(\mathbf{r}_t|\mathbf{p}_t, V_t) f(\mathbf{s}_t|\mathbf{p}_t, \mathbf{r}_t, V_t) \tag{5}$$

Splitting up the pose estimation in this way reduces the dimensionality of each sub-problem allowing fewer particles to be used during estimation.

### 2.3 Selection of Stable Landmarks

The set of stable landmarks is selected during training as follows. First, the annotation meshes are aligned to a common coordinate frame (see the next section). The correspondences formed during alignment identify each mesh vertex across all meshes (and volumes). The mesh vertices are used as landmark candidates. Denoting their location as $\{\mathbf{g}_i\}$. One position detector per landmark candidate is trained using all annotations. The detectors are then used to obtain detection results for each landmark, denoted as $\{\mathbf{d}_i\}$. The uncertainty of each detector is modeled by the covariance matrix, $\mathbf{C}_i$, of the final detected candidate location: $\mathbf{C}_i = \sum_i \mathbf{e}_i \mathbf{e}_i^\top$, where $\mathbf{e}_i = \mathbf{d}_i - \mathbf{g}_i$. The stable landmarks are selected one by one according to the score criterion $s_i = trace(\mathbf{C}_i)$ (higher $s$ indicates higher uncertainty). During this selection, we apply spatial filtering (with radius $r = 20mm$) using the score $s_i$. This way, we obtain a set of stable landmarks with low uncertainty that are widely distributed along the lung surface.

### 2.4 Shape Initialization

After the poses of left and right lung have been detected, the boundary of the lung is detected to find an initial segmentation. This initial segmentation is a deformation of a triangulated mesh model. The model, $\mathcal{M} = (\mathcal{P}, \mathcal{T})$ consists of a set of points, $\mathcal{P} = \{\mathbf{x}_i \in \mathbb{R}^3\}_{i=1}^N$, and a set of triangle indices, $\mathcal{T} = \{\triangle_j \in \mathbb{Z}^3\}_{j=1}^M$.

The high dimensional search space is restricted by a prior learned linear model of shape variation. The prior shape model, $\mathcal{S} = (\{\hat{\mathbf{x}}\}_{i=1}^N, \{U_j\}_{j=1}^M)$, consists of a mean shape and a set of linear basis shapes, $U_j = \{u_{ij}\}_{i=1}^N$, that are learned through procrustes analysis and PCA on training data. A synthesized shape in the *span* of the shape-space can be specified by a few PCA coefficients, $\{\lambda_j\}$, and a pose, $(\mathbf{p}, \mathbf{r}, \mathbf{s})$:

$$g(\mathbf{x}_i; \{\lambda_j\}, \mathbf{p}, \mathbf{r}, \mathbf{s}) = \mathbf{p} + \mathbf{M}(\mathbf{s}, \mathbf{r}) \sum_j (\hat{\mathbf{x}}_i + \mathbf{u}_{ij} \lambda_j) \tag{6}$$

where $\mathbf{M}(\mathbf{s}, \mathbf{r})$ is a $3 \times 3$ scale and rotation matrix.

Estimation of the first three coefficients is done in the HDN framework, where $\boldsymbol{\theta}_t = \{\lambda_1, \lambda_2, \lambda_3\}$. Particles from the pose estimation process are augmented

with three PCA coefficients sampled uniformly over the range of coefficients observed in the training data. Similar to Eq. 3, the observation model, $f(\boldsymbol{\theta}_t|V_t)$, is empirically modeled with a discriminative classifier that uses steerable features evaluated on surface points of the synthesized mesh [5].

### 2.5 Freeform Refinement

The first three PCA coefficients give a coarse approximation to the boundary. In order for the shape model to be expressive enough for all real instances, a large number of basis functions may be needed (e.g., the order of 100s). Instead of estimating all of the $\lambda$ coefficients directly as above, the freeform refinement takes an iterative surface deformation approach [5].

Starting with the initialized shape from above, the freeform refinement seeks to find the most probable mesh, $\mathcal{M}$, in the space of the linear shape model:

$$\max f(\mathcal{M}|V_t) \quad s.t. \quad \mathcal{M} \in span(\mathcal{S}) \tag{7}$$

Where $f(\mathcal{M}|V_t)$ is approximated by integrating over the surface:

$$f(\mathcal{M}|V_t) = \frac{1}{N}\sum_{x_i} f(\mathbf{x}_i|V_t). \tag{8}$$

Here the per-point posterior is directly approximated by a discriminative model. Letting $y_i = \{-1, +1\}$ be a random variable denoting the presence of a surface at point $\mathbf{x}_i$ along normal $\mathbf{n}_i$:

$$f(\mathbf{x}_i|V_t) = f(y_i = 1|\mathbf{x}_i, \mathbf{n}_i, V_t) \tag{9}$$

Instead of performing a coupled high dimensional optimization for all points simultaneously, local search within a predefined range $\{-\tau, \tau\}$ is performed for each vertex to find the best displacement along the normal, $\mathbf{x}_i \leftarrow \mathbf{x}_i + d_i\mathbf{n}_i$:

$$d_i = \arg\max_{-\tau \le d \le \tau} f(\mathbf{x}_i + d\mathbf{n}_i|V_t) \tag{10}$$

The resulting shape is projected onto the shape-space and surface normals are updated. This interleaved displace and regularization process is iterated several times. In latter iterations, $\tau$ is reduced, and the shape is allowed to vary from the $span(S)$. In these iterations, instead of regularizing by projecting into the shape space, a simple mesh smoothing is used to regularize the displaced mesh.

**Table 1.** Results of symmetrical point-to-mesh comparisons of detected results and annotations for both lungs, with and without stable landmark detection

| Lung | Landmark | Mean (std.) | Med. | Min | Max | 80% |
|---|---|---|---|---|---|---|
| right | no | 2.35 ± 0.86 | 2.16 | 1.40 | 6.43 | 2.57 |
| right | yes | 1.98 ± 0.62 | 1.82 | 1.37 | 4.87 | 2.18 |
| left | no | 2.31 ± 2.42 | 1.96 | 1.28 | 21.11 | 2.22 |
| left | yes | 1.92 ± 0.73 | 1.80 | 1.19 | 6.54 | 2.15 |

## 3 Experiments

Our experiments start by analyzing the error of detectors during training. We then show the set of top automatically selected landmarks. Finally, we provide qualitative and quantitative evaluation of lung segmentation.

Our dataset consists of 260 expert-annotated diagnostic CT scans of varying contrast. The slice thickness varies from 0.5 to 5.0. The dataset is randomly separated into two disjoint sets, one for training (192 volumes) and one for testing (68 volumes).

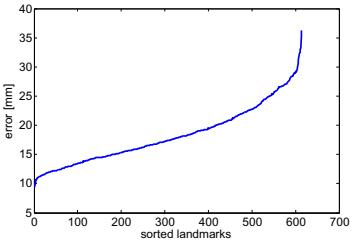


**Fig. 3.** Sorted mean errors of the 614 landmarks computed from all training volumes

The first result in Figure 3 shows the sorted errors of all candidate landmarks (Section 2.3). These are the detection results obtained from landmark detectors

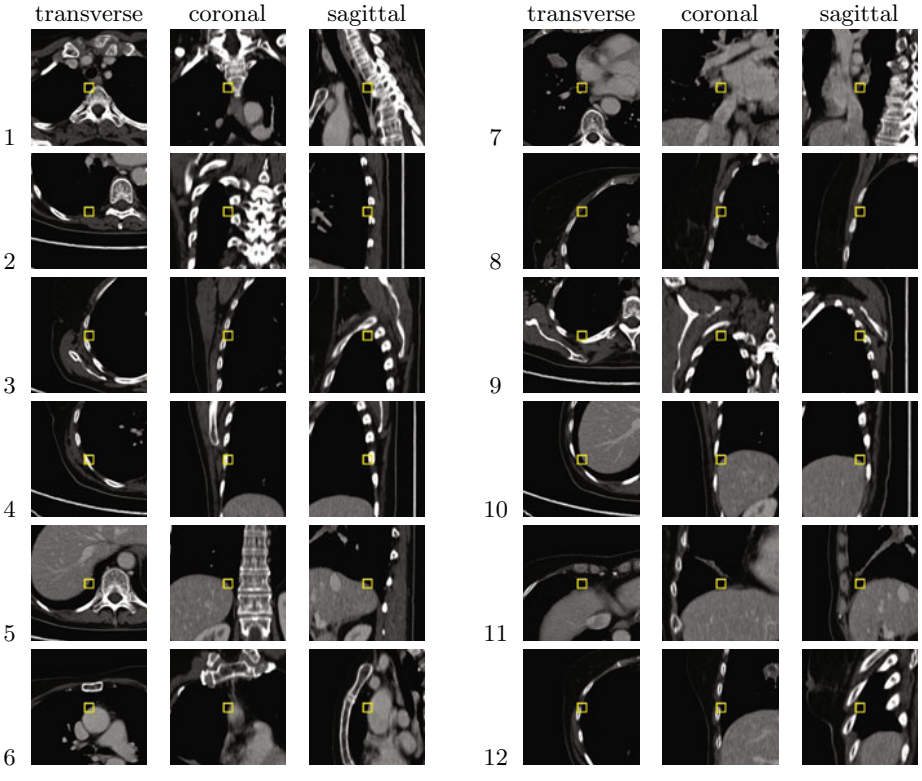


**Fig. 4.** Twelve strongest landmarks selected out of 614 with focus on spatial coverage. Notice that they are selected near distinctive anatomical structures such as ribs (3, 4, 5, 12), vertebrae (1, 2) and top (5) and bottom of the lung (9, 10, 11).

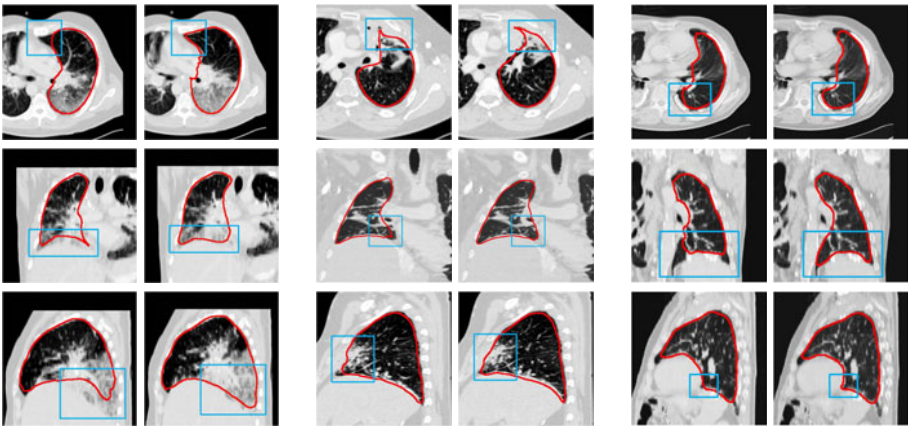

**Fig. 5.** Comparison of some results without stable landmarks (left column of every set) and with stable landmark detection (right column of every set)

trained using candidates formed from mesh vertices. The seemingly larger errors are caused by incorrect correspondences after mesh alignment. However, these landmarks are stable to provide accurate initialization for the mesh boundary refinement. We used 20 mm radius in landmark spatial filtering which resulted in 143 landmarks for the right lung and 133 landmarks for the left lung. To illustrate the effectiveness of the filtering, we set the radius to 70 mm to produce a set of 12 stable landmarks (Fig. 4). Notice that they are distributed across distinct locations inside the lung. Typically, landmarks near ribs and vertebrae are stable but not always. For example, landmark neighborhoods along some parts of the ribs or even across different ribs might not be distinctive enough.

Our final set of results analyzes performance of the lung segmentation. The algorithm was run as described in Section 2 and Figure 2. In the first experiment, the set of stable landmarks was used to initialize the boundary refinement. In the second experiment, the initialization was done by a mean mesh transformed according to the estimated pose. The errors summarized in Table 1 show that the initialization provided by stable landmarks helps to achieve significantly better accuracy ($p < 0.05$) of the final segmentations. The maximum error also decreased considerably and in the case of left lung one large failure was corrected. Several qualitative segmentation results involving pathologies are shown in Figure 5.

## 4 Conclusion

We proposed a robust learning-based technique for accurate lung segmentation in challenging CT volumes involving abnormalities. The technique first reliably detects the carina of the trachea as an anchor point for pose estimation of left and right lung. The poses are used to initialize a set of stable anatomical landmarks distributed on the lung surface. The stable landmarks are selected automatically

from a candidate set formed from vertices of mesh annotations by employing measures of uncertainty and spatial distribution. The initial landmark positions are refined and subsequently used to provide a rough estimate for the shape model and final lung boundary refinement.

We have shown the automatic landmark selection procedure determines a set of stable landmarks. These landmarks lead to improved initialization of the boundary refinement and ultimately higher accuracy of the final segmentations. Our future work focuses on further improvements especially near the lung sharp boundaries which are difficult to capture with a mesh representation.

## References

1. De Nunzio, G., Tommasi, E., Agrusti, A., Cataldo, R., De Mitri, I., Favetta, M., Maglio, S., Massafra, A., Quarta, M., Torsello, M., et al.: Automatic lung segmentation in CT images with accurate handling of the hilar region. Journal of Digital Imaging, 1–17 (2009)
2. Hu, S., Hoffman, E., Reinhardt, J.: Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. IEEE Trans. Med. Imag. 20(6), 490–498 (2002)
3. Kockelkorn, T., van Rikxoort, E., Grutters, J., van Ginneken, B.: Interactive lung segmentation in CT scans with severe abnormalities. In: IEEE Int. Symp. on Biomed. Imag., pp. 564–567 (2010)
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. In: IEEE Conf. on Comp. Vis. and Patt. Recog., Los Alamitos, CA, USA (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(9), 1173–1180 (2008)
7. Sluimer, I., Prokop, M., van Ginneken, B.: Toward automated segmentation of the pathological lung in CT. IEEE Trans. Med. Imag. 24(8), 1025–1038 (2005)
8. Sofka, M., Zhang, J., Zhou, S., Comaniciu, D.: Multiple object detection by sequential Monte Carlo and hierarchical detection network. In: IEEE Conf. on Comp. Vis. and Patt. Recog. (June 13-18, 2010)
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. Tu, Z.: Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering. In: IEEE Int. Conf. on Comp. Vis. and Patt. Recog., vol. 2, pp. 1589–1596 (2005)
11. Vidal, C., Hewitt, J., Davis, S., Younes, L., Jain, S., Jedynak, B.: Template registration with missing parts: Application to the segmentation of M. tuberculosis infected lungs. In: IEEE Int. Symp. on Biomed. Imag., pp. 718–721 (2009)
12. Wang, J., Li, F., Li, Q.: Automated segmentation of lungs with severe interstitial lung disease in CT. Medical Physics 36, 4592 (2009)