Model-Based Esophagus Segmentation from CT Scans Using a Spatial Probability Map

Johannes Feulner[1,3], S. Kevin Zhou[2], Martin Huber[3], Alexander Cavallaro[4], Joachim Hornegger[1], and Dorin Comaniciu[2]

[1] Pattern Recognition Lab, University of Erlangen-Nuremberg, Germany
[2] Siemens Corporate Research, Princeton, NJ, USA
[3] Siemens Corporate Technology, Erlangen, Germany
[4] Radiology Institute, University Hospital Erlangen, Germany

Abstract. Automatic segmentation of the esophagus from CT data is a challenging problem. Its wall consists of muscle tissue, which has low contrast in CT. Sometimes it is filled with air or remains of orally given contrast agent. While air holes are a clear hint to a human when searching for the esophagus, we found that they are rather distracting to discriminative models of the appearance because of their similarity to the trachea and to lung tissue. However, air inside the respiratory organs can be segmented easily. In this paper, we propose to combine a model based segmentation algorithm of the esophagus with a spatial probability map generated from detected air. Threefold cross-validation on 144 datasets showed that this probability map, combined with a technique that puts more focus on hard cases, increases accuracy by 22%. In contrast to prior work, our method is not only automatic on a manually selected region of interest, but on a whole thoracic CT scan, while our mean segmentation error of 1.80mm is even better.

1 Introduction

Atrial fibrillation is a major cause of stroke. It can be treated with a catheter ablation therapy in the heart. However, this intervention imposes the risk of an atrio-esophageal fistula. The air from the esophagus can enter the left atrium, which normally causes the death of the patient [1]. Here, a segmentation of the esophagus can be helpful during intervention planning. A segmentation can also help to find lymph nodes in CT images of the chest area. The attenuation coefficients are so similar that it is often impossible even for a human to separate them given a single slice. Here, a segmentation can provide valuable overview to a physician, and also automatic detectors benefit because the esophagus can be excluded from search.

Automatic segmentation of the esophagus is challenging because it easily can be confused with vessels, muscles or lymph nodes. Both shape and appearance can vary a lot. It may appear solid, but it can also contain air or remains of contrast agent given orally to the patient.

Prior work on the topic is limited. In [2], the esophagus is segmented with a non-parametric shape prior of the centerline and an appearance model.

T. Jiang et al. (Eds.): MICCAI 2010, Part I, LNCS 6361, pp. 95–102, 2010.

The method is semi-automated: The user has to specify two points on the centerline. Moreover, segmentations of the left atrium and the aorta are required as anchors for the shape prior. Another semiautomatic segmentation method which also relies on a shape prior that is relative to surrounding structures is described in [3]. In [4], contours that were manually drawn into axial slices are interpolated without using the image itself.

In this work, we follow the approach of [5]. There, a four step method is proposed. First, a box detector based on a discriminative model is run for each axial slice and a set of candidates of the approximate esophagus contour is generated. Then, the candidates are clustered and merged. Now a Markov chain model is used to infer the path through the axial slices that has the highest probability. Finally, a surface is generated and refined, again with a discriminative model. However, the method is only automatic on a manually cropped region of interest (ROI) and not on a whole CT volume.

In this work, the method of [5] is extended in three ways. First, it is made fully automatic on uncropped CT volumes as acquired by the scanner. The region of interest is determined by detecting a salient landmark in the chest. Second, the accuracy was improved by incorporating a new intermediate detection step. Finally, we explicitly handle air in the esophagus and in the respiratory organs in order to further increase the robustness.

The remainder of the paper is structured as follows: In section 2.1, we shortly summarize the approach of [5]. Section 2.2 describes our approach for automatically detecting the region of interest. In section 2.3, the detector which finds esophagus contour candidates is described along with the new intermediate detection step. Section 2.4 explains how the distribution of air was incorporated. Section 3 presents experiments and results, and section 4 concludes the paper.

2 Method

2.1 Model Based Segmentation

In [5], the esophagus contour in axial slices is first approximated with ellipses. They are parameterized as

$$\boldsymbol{e} = (\boldsymbol{t}, \theta, \boldsymbol{s}), \tag{1}$$

where $\boldsymbol{t} = (x, y)$ is the center within the slice, θ is the rotation angle and $\boldsymbol{s} = (a, b)$ contains the semi-axes of the ellipse.

Ellipses are detected using a technique called marginal space learning [6]. Instead of directly searching the five dimensional search space, three classifiers are trained. As classifiers, we use probabilistic boosting-trees (PBT). These are binary decision trees with a strong AdaBoost classifier at each node. The first PBT classifier learns the probability $p(m = 1|\boldsymbol{H}(\boldsymbol{t}))$ of whether there is a model instance in the image at a certain location $\boldsymbol{t}$. Here, $\boldsymbol{H}(\boldsymbol{t})$ denote Haar-like features extracted at position $\boldsymbol{t}$. These are simple cuboid filters similar to the rectangle filters described in [7]. They can be computed very efficiently and thus even allow to search the volume exhaustively. The second one is trained to learn

the probability $p(m = 1|\boldsymbol{S}(\boldsymbol{t}, \theta))$ of whether there is a model instance at a given position with a given rotation angle θ. It uses steerable features $\boldsymbol{S}$ as proposed in [6], which are simple point features sampled on a regular grid, and the sampling pattern is rotated according to θ. The third classifier is trained on the whole search space to learn $p(m = 1|\boldsymbol{S}(\boldsymbol{t}, \theta, \boldsymbol{s}))$. It uses steerable features as well, but now the sampling pattern is also scaled. By first using the detectors trained on marginal spaces, large portions of the search space can be pruned at early stages. The result is a set of ellipse candidates $\{\boldsymbol{e}^{(1)} \dots \boldsymbol{e}^{(N)}\}$ per axial slice. These candidates are spatially clustered and merged. Each cluster center $\boldsymbol{c}^{(k)}$ is associated with a weight $\sigma^{(k)}$ which is the sum of detection scores $p(m = 1|\boldsymbol{e}^{(i)})$ of candidates belonging to cluster k.

Now the most likely path through the axial slices is inferred using a Markov chain model. Each slice i is associated with a random variable $\boldsymbol{s}_i$. Possible states of the random variable $\boldsymbol{s}_i$ are the cluster centers $\boldsymbol{c}^{(k)}$ of slice i. The transition probability $p(\boldsymbol{s}_{i+1}|\boldsymbol{s}_i)$ from one slice to the next is modelled by normal distributions whose parameters are estimated from manual annotations, and the probability $p(\boldsymbol{s}_i|\boldsymbol{v})$ of a variable given an observed image slice $\boldsymbol{v}$ is set to weight $\sigma^{(i)}$. The maximum a posteriori estimate of all state variables given all observed image slices $p(\boldsymbol{s}_{1:T}|\boldsymbol{v}_{1:T})$ can be computed efficiently using dynamic programming.

After the best path has been computed, it is converted into a triangulated surface representation and refined to better fit the boundary of the organ. The vertices of the surface are individually deformed according to the output of another PBT classifier which was trained with steerable features. Then a mesh smoothing algorithm is applied. Deformation and smoothing is repeated once.

2.2 Automatic ROI Detection

The method described in section 2.1 only works on a region of interest showing the esophagus. In order to make it fully automatic also on uncropped CT scans showing for instance the thoracic and abdominal region, we detect the ROI automatically. A salient point close to the esophagus that can be detected very robustly is the bifurcation of the trachea. The landmark detection method used here is described in [8]. An axis-aligned cuboidal region is rigidly attached to that landmark such that the esophagus is always inside in 144 datasets with a minimum margin of 3cm in x and y direction. The resulting region is of size $13.3 \times 15.6\text{cm}^2$. In vertical (z) direction, the size is set to 26cm. This ROI is fairly large, which makes the detection harder because it contains more structures that may be confused with the esophagus.

2.3 Discriminative Model

Within the detected ROI, we run for each axial slice a classifier that was trained to learn $p(m = 1|\boldsymbol{H}(\boldsymbol{t}))$ as described in section 2.1. A set of candidates $C_{T1} = \{\boldsymbol{t}_1 \dots \boldsymbol{t}_{N_{T1}}\}$ is generated which contains the N_{T1} positions with highest detection score. We now propose to train another classifier of the same type and the same features. But now, the negative training examples are generated by

scanning images with the first classifier and collecting false alarms instead of randomly drawing samples from the images. Thus, the second classifier gets specialized on the difficult cases. It only considers the candidates from C_{T1} and generates a set C_{T2} which contains the N_{T2} candidates with highest scores of the second detector. The remaining steps in the detection pipeline are similar to [5]: A classifier trained to learn the probability $p(m = 1|\boldsymbol{S}(\boldsymbol{t}, \theta))$ considers the candidate set C_{T2} and generates a set C_{TR} of position and orientation candidates of size N_{TR}. Finally, a classifier that was trained on the full search space of translation, rotation and scale is run on C_{TR} to generate the ellipse candidates of a slice. Clustering, merging, path inference and surface generation are adopted from [5].

In Fig. 1, example output is displayed for each step of the detection pipeline.

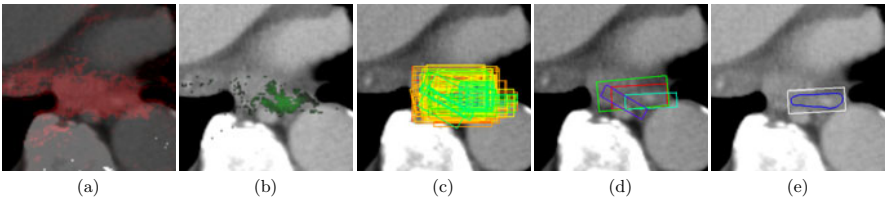


Fig. 1. The proposed detection pipeline. (a) shows the detection scores generated by the first translation detector. The output of the second translation detector is visualized in (b). The bounding boxes of the ellipse candidates are shown in (c). The confidence is color coded in HSV color space: Violet means lowest score, red is highest. The candidate set is clustered and merged (d). The best path through the axial slices is inferred, converted into a surface and further refined (e). The blue contour is the final segmentation.

2.4 Including the Distribution of Air

When a human tries to find the esophagus in a CT dataset, s/he also looks for air holes as they are clearly visible and a strong indicator for the esophagus. One might expect the classifiers to learn this correlation, but we found that air holes rather distract them. The reason is that the classifiers only rely on local features. Then, air holes can be confused easily with the trachea or lung tissue, and both are a priori much more likely because they cover a larger volume. A human, however, recognizes and excludes the respiratory organs effortlessly.

Separating esophageal from respiratory air. Therefore, we propose to support the detector by adding the knowledge that air belonging to the respiratory organs cannot belong to the esophagus, and air elsewhere most likely is inside the esophagus. This is modelled with a binary mask $B(\boldsymbol{t})$ that is zero if $\boldsymbol{t}$ belongs to a respiratory organ and one otherwise, and a probability map $S(\boldsymbol{t})$ of the esophagus based on detected air holes.

Segmenting the air of the respiratory organs in CT is straightforward because it is one connected region. First, voxels with an attenuation coefficient below -625HU are labeled as air. Vessels and airways in the lung are labeled as air as well by identifying small connected components in axial, sagittal and coronal slices. Now all regions marked as air which touch the left, right, front or back border of the ROI are removed. The result is stored in B. Remaining regions marked as air probably belong to the esophagus. They are labeled with 1 in a mask E. Elsewhere, E is zero. A similar method to find esophageal air holes is described in [9]. Now for each axial slice, it is checked whether E contains exactly one connected region labeled as esophageal air. If so, we set the current slice of the probability map S to $g(|\boldsymbol{t} - \boldsymbol{p}|)$, where $\boldsymbol{p}$ is the point of gravity of the region within the slice and g is a Gaussian with standard deviation s that is deformed and trimmed to have a maximum of 1 and limited support in $[-w, w]$. We selected a value of 7mm as s and 10mm as w.

Integration into detector. We now define a combined probability map $C(\boldsymbol{t})$ as

$$C(\boldsymbol{t}) = \frac{B(\boldsymbol{t}) + S(\boldsymbol{t})}{2} \tag{2}$$

and model the probability $p(m = 1|C(\boldsymbol{t}))$ of observing the esophagus at position $\boldsymbol{t}$ given the global distribution of air as being proportional to $C(\boldsymbol{t})$:

$$p(m = 1|C(\boldsymbol{t})) \propto C(\boldsymbol{t}). \tag{3}$$

During position detection, we are finally interested in the probability $p(m = 1|\boldsymbol{H}(\boldsymbol{t}), C(\boldsymbol{t}))$ of observing the esophagus at a certain location $\boldsymbol{t}$ given the Haar-like feature response $\boldsymbol{H}(\boldsymbol{t})$ and the information from the global distribution of air $C(\boldsymbol{t})$. In order to simplify the notation, we will omit the argument $\boldsymbol{t}$ in the remainder of this section. Using Bayes' rule, this can be rewritten as

$$p(m = 1|\boldsymbol{H}, C) = \frac{p(\boldsymbol{H}, C|m = 1)p(m = 1)}{p(\boldsymbol{H}, C)} \tag{4}$$

Now we assume that the feature vector $\boldsymbol{H}$ is statistically independent from the distribution of air C. This is of course a simplifying assumption. $\boldsymbol{H}$ and C are to some extent statistically dependent. The assumption is justified by the fact that the map C does improve the performance as we will see, which means that $\boldsymbol{H}$ does not contain much information about C. With this assumption, (4) can be transformed into

$$p(m = 1|\boldsymbol{H}, C) = \frac{p(\boldsymbol{H}|m = 1)p(C|m = 1)p(m = 1)}{p(\boldsymbol{H})p(C)} \tag{5}$$

$$= \frac{p(m = 1|\boldsymbol{H})p(m = 1|C)}{p(m = 1)}, \tag{6}$$

which is proportional to the product $p(m = 1|\boldsymbol{H})C(\boldsymbol{t})$ of the classifier output and the probability map C. This means we can integrate C into the translation detector simply by multiplying it with the detection score. In Fig. 2, the probability map C is visualized for two axial CT slices.

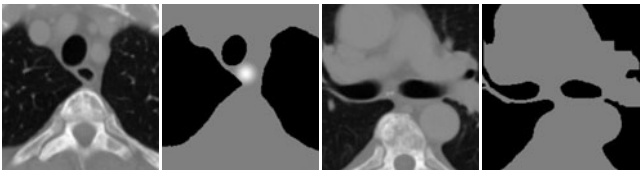


Fig. 2. Two examples of CT slices along with their combined probability map $C(\boldsymbol{t})$ generated from the distribution of air inside the volume. Left: The air hole is a clear hint for the esophagus. Right: No air hole is present, but respiratory air can be excluded.

Regions filled with respiratory air are not considered by the detector. Therefore, we also do not generate negative training examples from these regions. This makes the learning problem easier because now air is a priori more likely to be part of the esophagus.

3 Results

The method has been evaluated on 144 CT scans of the thoracic or the thoracic and abdominal region. No patient was included twice. The voxel spacing in x and y direction was in the range of 0.7mm to 0.8mm. The spacing in (longitudinal) z direction was 5mm. After ROI detection, the volumes were resampled to a voxel spacing of $0.7 \times 0.7 \times 5\text{mm}^3$. Manual segmentations were available for all datasets. The segmentations typically ranged from the thyroid gland down to a level below the left atrium.

The accuracy was measured using threefold cross-validation. For each fold, all five classifiers for translation ($2\times$), orientation, scale and surface were trained on the training data, and the parameters of the Markov model were estimated from the same training data. The remaining data was used for testing. There was no overlap between training and testing data. For evaluation, the detector was run in z direction on the same interval covered by the manual annotation in order not to introduce artificial errors because of different lengths of the segmentations.

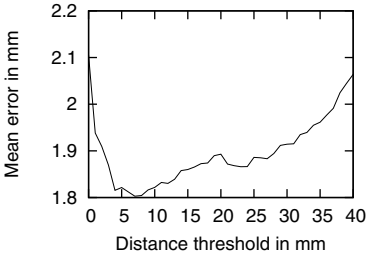
ROI detection succeeded in all of the 144 datasets. Due to the large ROI, the segmentation method can tolerate normal anatomical variations and detection errors of the bifurcation of the trachea.

Table 1 shows the results of accuracy evaluation. As error measures, we used the mean segmentation error and the Hausdorff distance, averaged over all test datasets. The mean error measures the mean distance between the ground truth and the detected surface, while the Hausdorff distance measures the maximum. The number of candidates N_{T1} generated by the first translation detector was set to 400, N_{T2} to 120, N_{TR} to 50, and the number of final candidates per slice to 200. The distance threshold in the clustering step was set to 8mm.

The proposed method segmented the esophagus with a mean error of 1.80mm. If only the binary mask B is used instead of the combined probability map C, the error is 1.88mm. If the air model as described in section 2.4 is omitted, the error is 1.94mm, meaning that explicitly modelling the air significantly improved the

Table 1. Results of performance evaluation. Shown is the mean error and the mean Hausdorff distance along with the corresponding standard deviations.

Method	mean error in mm	Hausdorff dist. in mm
Proposed method	1.80 ± 1.17	12.62 ± 7.01
Only binary air model $B(\boldsymbol{t})$	1.88 ± 1.24	13.00 ± 7.88
No air model	1.94 ± 1.39	13.06 ± 7.21
Single translation class.	2.07 ± 1.47	14.50 ± 8.92
No air model, single translation class.	2.32 ± 1.87	15.02 ± 9.83
No Markov model	2.30 ± 1.49	17.29 ± 11.42
Proposed method, best 80%	1.34 ± 0.31	9.65 ± 3.07
Feulner et al. [5]	2.28 ± 1.58	14.5
Inter observer variability	0.78 ± 0.17	7.29 ± 2.22



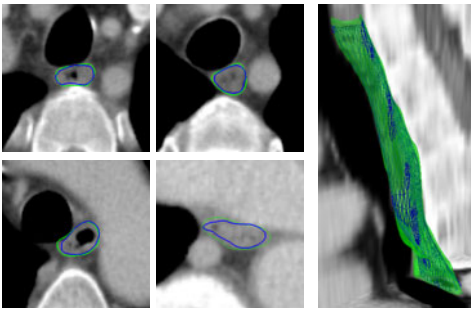


Fig. 3. Left: The mean segmentation error for different distance thresholds used for clustering. Right: Examples of automatic segmentations on unseen data (blue) along with ground truth (green).

accuracy. If the air probability map is used, but the second translation detector as described in section 2.3 is omitted, we get an error of 2.07mm. Without both the air model and the additional detector, performance is 2.32mm, meaning that using both decreases the error by 22%. When the Markov model is turned off and for each slice, the ellipse candidates are simply averaged, the error becomes 2.30mm. This is different from what is described in [5]: There, the use of the Markov chain did hardly influence the numeric results. Here, it clearly improves the accuracy. The reason is that the Markov model is especially useful to resolve ambiguity that occurs much more frequently in a larger ROI. The data used for evaluation does also contain extreme cases which in principle can be handled by our method if enough training examples are available. This was not always the case. If the most difficult cases are excluded from the test set, the mean error becomes 1.34mm. For comparison, the results of [5] are shown. We furthermore did an experiment to measure the inter observer variability: Ten datasets were manually segmented a second time by another observer. The second segmentations were evaluated like automatic ones. The result is shown in the last row.

Table 2. Computation time in seconds for different steps of the method

ROI detec.	prob. map gen.	ellipse detec.	path inference	refinement	**total**
6.96	1.13	7.40	$0.40 \cdot 10^{-3}$	0.34	**15.83**

Table 2 shows the computational requirements of the different steps of the method, measured on a 2.2GHz dual core PC. Though the ROI is larger, ellipse detection is slightly faster compared to [5] due to the spatial probability map and the second translation detector which rejects most candidates of the first one. In total, the method takes less than 16s.

4 Discussion

In contrast to prior work, our method is able to segment the esophagus from uncropped CT volume images without any user interaction. Segmentation on an automatically detected ROI is harder because the ROI has to be made large. Still, our results are better than what has been reported in [5]. A mean error of 1.8mm, which is only 1.0mm above the inter observer variability, can be considered as a good result for this problem.

References

1. Pappone, C., Oral, H., Santinelli, V., Vicedomini, G., Lang, C.C., Manguso, F., Torracca, L., Benussi, S., Alfieri, O., Hong, R., Lau, W., Hirata, K., Shikuma, N., Hall, B., Morady, F.: Atrio-Esophageal Fistula as a Complication of Percutaneous Transcatheter Ablation of Atrial Fibrillation. Circulation 109(22), 2724–2726 (2004)
2. Rousson, M., Bai, Y., Xu, C., Sauer, F.: Probabilistic minimal path for automated esophagus segmentation. In: Proceedings of the SPIE, vol. 6144, pp. 1361–1369 (2006)
3. Kurugol, S., Sharp, G., Dy, J., Brooks, D.: Esophagus Segmentation in Thoracic CT Images for Radiotherapy Planning. In: AAPM, vol. 36, pp. 2454–2454 (2009)
4. Fieselmann, A., Lautenschläger, S., Deinzer, F., John, M., Poppe, B.: Esophagus Segmentation by Spatially-Constrained Shape Interpolation. In: BVM, p. 247–251 (2008)
5. Feulner, J., Zhou, S.K., Cavallaro, A., Seifert, S., Hornegger, J., Comaniciu, D.: Fast Automatic Segmentation of the Esophagus from 3D CT Data Using a Probabilistic Model. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 255–262. Springer, Heidelberg (2009)
6. 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 (2007)
7. Viola, P., Jones, M.: Rapid Object Detection using a Boosted Cascade of Simple Features. In: CVPR, vol. 1, p. 511 (2001)
8. Seifert, S., Barbu, A., Zhou, S.K., Liu, D., Feulner, J., Huber, M., Suehling, M., Cavallaro, A., Comaniciu, D.: Hierarchical parsing and semantic navigation of full body CT data. In: Medical Imaging, vol. 7259, p. 725902. SPIE (2009)
9. Fieselmann, A., Lautenschläger, S., Deinzer, F., Poppe, B.: Automatic Detection of Air Holes Inside the Esophagus in CT Images. In: BVM, p. 397-401 (2008)