# Supervised Probabilistic Segmentation
# of Pulmonary Nodules in CT Scans

Bram van Ginneken

Image Sciences Institute, University Medical Center Utrecht, the Netherlands
`bram@isi.uu.nl`

**Abstract.** An automatic method for lung nodule segmentation from computed tomography (CT) data is presented that is different from previous work in several respects. Firstly, it is supervised; it learns how to obtain a reliable segmentation from examples in a training phase. Secondly, the method provides a soft, or probabilistic segmentation, thus taking into account the uncertainty inherent in this segmentation task. The method is trained and tested on a public data set of 23 nodules for which soft labelings are available. The new method is shown to outperform a previously published conventional method. By merely changing the training data, non-solid nodules can also be segmented.

## 1   Introduction

Lung cancer is by far the main cancer killer. Early detection of malignant nodules from computed tomography scans and subsequent surgery, is considered one of the most promising strategies to reduce lung cancer mortality. Only a small fraction of nodules is cancerous and volume and growth rate of the lesions are the main indicators of malignancy. Thus accurate segmentation of small pulmonary nodules is of paramount clinical importance. Many academic groups and companies have developed lung nodule segmentation schemes, see for example Refs. [1,2,3,4,5,6,7] and [8] for a more extensive overview. Most of these approaches are based on standard image processing techniques such as thresholding, region growing, component labeling, ray shooting, mathematical morphology, etc. Many are semi-automatic, requiring internal parameters to be tuned for every specific case.

A major impediment for algorithm development and validation is the impossibility to obtain a ground truth for nodule segmentations in real clinical data. Most algorithms have therefore been validated on phantom data, with good results. But these phantoms do not capture the enormous variety of pulmonary structures and scan artifacts encountered in real data. Our experience with several commercial algorithms from various vendors is that they regularly produce unsatisfactory segmentations, even on simple cases of isolated nodules. Reproducibility studies have shown [9] that 95% limits of agreement for volume measurements are around 20% for commercial software with manual interaction by a radiologist. This limits the possibility to unequivocally demonstrate

growth of nodules on short-term follow-up scans. Hence the problem of nodule segmentation is still far from solved.

Recently the Lung Image Database Consortium (LIDC) [10], published a set of 23 scans with manual nodule outlines. Analysis of this data, that is used in this work, reveals that only 15% of voxels with probability for being nodule $p_i > 0$ have $p_i = 1$ and only an additional 18% has $p_i \geq 0.9$. This suggests that nodule segmentation is not a *hard* classification task and that it is preferably to perform *probabilistic* or *soft* segmentations, contrary to what published methods do.

In this work we propose a new method for probabilistic nodule segmentation that takes the uncertainty about the voxel label explicitly into account. We use the LIDC data for training, and employ statistical techniques to infer the probability that voxels in unseen data belong to a nodule. The method is completely automatic if a seed point in the nodule is supplied, and can be applied to different types of nodules by only changing the training data.

## 2    Materials

The LIDC data consists of 23 scans in which one nodule has been manually segmented by 6 radiologists using manual tracing and two (unspecified) semi-automatic algorithms, making for a total of 18 manual segmentations. Together with CT data, soft segmentations are available, derived by averaging the 18 hard segmentations. These soft segmentations are taken as truth in this work. Slice thickness is 0.625 mm, in-plane resolution varies from 0.625 to 0.742 making for nearly isotropic data. All volumes and length parameters in this work are given in voxels. The variety among the nodules is large. They range from very small to very large, from round and solid to spiculated, ill-defined and non-solid.

## 3    Method

The method assumes that we are supplied with a CT scan and a seed point in the nodule to be segmented. The result is independent of the exact position of the seed point as long as the seed is inside the nodule. As a pre-processing step, the lung fields in the scan are segmented. This allows for correct segmentation of nodules in contact with the chest wall. The data does not consist of complete scans but only of sections around the nodule. Therefore a 2D lung field segmentation algorithm was used, similar to the one described in [11] and based on thresholding, component-labeling, rule-based selection of lung field components, hole filling and smoothing the lung fields by a morphological closing operation with a circular kernel. The results were satisfactory in all cases. All subsequent analysis is performed on voxels within the lung fields only.

### 3.1    Region Regression

The core of the soft segmentation algorithm consists of a region growing process starting from the supplied seed point. We used 6-connectivity. For every neighbor

encountered during the growing process, the probability $p_i$ for it belonging to the nodule is estimated using regression. All neighbors with $p_i > t$ are accepted. We used $t = 0.15$ but slight changes to this value are not critical.

Regression is a general statistical technique to relate one or more predictor variables (features, that form a feature vector that resides in feature space) to a single response variable (output value, $p_i$). The functional relationship between the input and the output values is learned from examples and can then be used to predict the output of a new, previously unseen feature vector. We experimented with linear regression and two types of non-linear regression, $k$-nearest neighbor regression ($k$NNR) and support vector machine regression.

Both nonlinear methods gave similar results, better than linear regression. As $k$NNR was fastest and easiest to use, we report here results for this method only. The number of neighbors $k$ was set to 15 (again, this setting was not critical). The average output of these neighbors is the output of $k$NNR.

### 3.2    Features

A proper choice of features is crucially important to obtain good performance. Density of voxels is the most obvious feature to use, as most lung nodules have tissue densities which distinguish them from the surrounding lung parenchyma. To incorporate gray level information from a spatial neighborhood we used the original density as well as five blurred versions of the volume (Gaussian blurring with $\sigma = 0.5, 1, 2, 4, 8$) as features. To separate voxels on edges from those in homogeneous regions in features space, we added the gradient magnitude, computed with Gaussian derivatives at the same scales.

The main reason why simple region growing only produces good results in the case of isolated nodules is that other structures with tissue density, mainly vessels, are frequently in contact with the nodule. The remaining features aim to distinguish nodule voxels from those in attached structures. Vessels are elongated and a classic way to distinguish nodular from elongated structures is to examine second order derivatives along principal directions. We computed the eigenvalues of the Hessian matrix, sorted from large to small, again at scales $\sigma = 0.5, 1, 2, 4, 8$. Another way to remove dense structures that have limited extent in any direction (like vessels, or fissures) is by gray level openings. We performed a separable gray level opening with a uniform kernel of 3 voxels length. This process was repeated four times, yielding four feature values. Finally, we applied an algorithm described as Iterative Morphological Filtering (IMF) in [2]. This is a sequence of morphological operations aimed at removing small vessels that are in contact with a nodule. First a volume of interest around the nodule is thresholded at threshold $T$ and stored as $I$. Next, a binary opening with a spherical kernel of diameter $D$ is applied to $I$ and only the component connected to the seed point in the nodule is retained. This result is iteratively dilated with a spherical kernel of size $D$, $D/2$, $D/4$ and so on until $D = 1$. After every dilation, the result is multiplied with $I$. With only the opening applied, the vessels are removed but the shape of the nodule boundary changes substantially. The iterated dilation and multiplication process restores as much as possible of the original nodule.

We applied IMF with $T = $ -550 HU and $D = 11$ voxels and used the result, as well as five blurred versions ($\sigma = 0.5, 1, 2, 4, 8$) as feature volumes.

A set of feature vectors from voxels with known output $p_i$ is needed to train the regression system. To sample a balanced learning set from the nodules and their surroundings, 1000 voxels from every training case were randomly selected for which $p_i > 0$ and another 1000 were chosen among those with $p_i = 0$ and the distance to the closest voxel with $p_i > 0$ below 10 voxels.

### 3.3   Nonsolid Nodules

For nonsolid nodules, IMF is not effective, not even with different settings of $T$ and $D$. Also the other features are very different for these nodules. Solid nodules have fairly constant (low gradient) densities around 0 HU, while non-solid nodules have fluctuating (high gradient) densities around -500 HU. Training a regression system with predominantly solid nodules and applying it to non-solid nodules is therefore bound to give poor results. We developed a simple algorithm that inspects the density around the seed point, and decides if a nodule is solid or non-solid. We then trained different regression systems for solid and non-solid lesions. In the case of non-solid nodules, IMF features were not used.

### 3.4   Conventional Method

As reference method a published lung nodule segmentation algorithm [2] was implemented, which we believe representative of most existing methods. This method is based on the IMF algorithm. A volume of interest (VOI) around the nodule and a seed point is required. The VOI is first supersampled to 0.25 mm isotropic voxel size using tri-linear interpolation. Next, the IMF algorithm is applied. Finally, the VOI is subsampled to the original resolution by voxel averaging. Note that in this way a soft segmentation is obtained, but voxels with $0 < p_i < 1$ can only occur around the nodule boundary. We determined $T = -550$ HU and $D = 31$ voxels (in the supersampled data) as best parameter settings.

## 4   Experiments and Results

A soft or probabilistic segmentation $s$ is a volume in which each voxel has a probability $0 \leq p_i \leq 1$ for belonging to the object associated with it. To compute object volume from $s$ we use

$$SV(s) = \sum_{i=1}^{n} p_i,\tag{1}$$

where the summation runs over all voxels.

To compare two soft segmentations we define the soft overlap as

$$SO(s_1, s_2) = \sum_{i=1}^{n} \frac{\min(p_{1i}, p_{2i})}{\text{abs}(p_{1i} - p_{2i}) + \min(p_{1i}, p_{2i})}.\tag{2}$$

This definition is a straightforward extension to the soft case of the common definition of overlap for hard segmentations as intersection divided by union. It is equal to 1 for complete agreement and has a minimum value of 0 for complete disagreement.

Any soft segmentation $s$ can be turned into a hard (binary) segmentation $h$ by thresholding at 0.5. $SO(s,h)$ is a measure of the 'softness' of a segmentation, it is the best possible soft overlap with a given $s$ that could be obtained with a hard segmentation.

We report the soft volume of each nodule and $SO(s,h)$. The latter can be compared to $SO$ for both the new and the conventional method. We also report the relative (soft) volume error $RVE$ in % which is the relative difference (positive for oversegmentation and negative for undersegmentation) in volume between the true and the automatically determined soft volume.

As the number of nodules cases released by the LIDC is fairly small, all experiments were carried out in a leave-one-out regime. That means that for testing one case, all other cases are used for training. The number of non-solid cases turned out to be three. So there are 19 cases available for training for every solid nodule and only 2 for every non-solid nodule.

The results of the experiments are given in Table 1 and illustrated for a number of cases in Figure 1. The overlap values are significantly higher for the proposed method than for the conventional method which fails completely on all non-solid nodules and on case 1 (attached fissure is included in the segmentation), and case 3 (complex vascular attachment). For case 1 the new method provides an excellent segmentation. Case 3 is the only large failure of the proposed method, probably because there is no other case like this one available for training. Note how both methods produce good results for a large solid, almost isolated nodule (case 15) and even for a complex nodules such as case 10 (although the soft volume obtained with the proposed method agrees much better with the truth than the result of the conventional method). The overlap obtained with the proposed method is in 17 cases higher than the maximum possible overlap of a hard segmentation (columns $SO(s,h)$ vs. $SO$ in Table 1). The relative volume errors are fairly large but the proposed method achieves an $RVE < 15\%$ in 14 cases.

## 5   Discussion

The results clearly demonstrate the possibility of obtaining a fairly reliable soft nodule segmentation with a simple, voxel-wise methodology. The complete segmentation requires in the order of 20 to 30 seconds, including computation of feature images. There is ample room for improving the speed of the method. The differences between the proposed method and the conventional method are in many cases not big, which is understandable as the result of the IMF algorithm was used as features for the regression method.

The results for non-solid nodules are encouraging, taking into account that just two nodules were used for training the system in these cases. Using a

**Table 1.** Experimental results for all 23 nodule cases. The last two columns show the results of the conventional method. Non-solid nodules are indicated with an asterix.

| case # | $SV$ | $SO(s,h)$ | $SO$ | $RVE$ | $SO$ [2] | $RVE$ [2] |
|---|---|---|---|---|---|---|
| 1 | 772.2 | 0.74 | 0.89 | 3.7 | 0.28 | 253.7 |
| 2 | 215.3 | 0.65 | 0.80 | 9.4 | 0.67 | 37.8 |
| 3 | 121.3 | 0.59 | 0.13 | 586.0 | 0.10 | 842.0 |
| 4 | 243.9 | 0.64 | 0.68 | 0.4 | 0.49 | 88.0 |
| 5 | 307.2 | 0.61 | 0.80 | -1.1 | 0.67 | 22.0 |
| 6 | 144.2 | 0.59 | 0.73 | 27.4 | 0.68 | 34.0 |
| 7* | 347.0 | 0.63 | 0.52 | -44.6 | 0.01 | 7456.6 |
| 8 | 251.3 | 0.64 | 0.79 | 10.0 | 0.70 | 27.7 |
| 9 | 1171.0 | 0.66 | 0.67 | 11.5 | 0.58 | 67.5 |
| 10 | 21653.3 | 0.72 | 0.78 | 7.7 | 0.68 | 34.4 |
| 11 | 150.4 | 0.56 | 0.65 | 20.4 | 0.63 | 43.1 |
| 12 | 385.1 | 0.55 | 0.70 | 25.9 | 0.59 | 7.6 |
| 13 | 67.7 | 0.56 | 0.47 | -52.4 | 0.50 | -45.7 |
| 14 | 138.3 | 0.59 | 0.77 | -10.2 | 0.68 | -5.4 |
| 15 | 8125.5 | 0.84 | 0.90 | 0.2 | 0.85 | 15.6 |
| 16 | 541.8 | 0.41 | 0.54 | -39.7 | 0.37 | -48.8 |
| 17 | 113.1 | 0.54 | 0.70 | 7.8 | 0.64 | 4.3 |
| 18* | 912.4 | 0.53 | 0.29 | 6.9 | 0.04 | 1845.5 |
| 19 | 4472.6 | 0.55 | 0.51 | -32.9 | 0.56 | 4.2 |
| 20 | 464.7 | 0.67 | 0.79 | -9.5 | 0.71 | 29.8 |
| 21* | 623.6 | 0.46 | 0.60 | 23.4 | 0.08 | 977.7 |
| 22 | 1515.9 | 0.69 | 0.74 | 4.4 | 0.64 | 40.7 |
| 23 | 19497.3 | 0.79 | 0.77 | -9.1 | 0.78 | 15.8 |
| mean | 2705.9 | 0.62 | 0.66 | 23.7 | 0.52 | 510.8 |
| std | 5928.6 | 0.10 | 0.18 | 124.5 | 0.25 | 1577.7 |

larger training database is clearly a priority. Many of the current 23 cases are unique and therefore not represented in the training data in a leave-one-out test regime. There is no doubt that this has a detrimental effect on the segmentation performance.

Obvious issues for further research are investigation of other feature sets, different regression schemes and the use of feature selection and extraction methods. More interesting is the extension to a method that employs a second region regression based on features derived from the output of the first process. In this way features that are more directly related to the shape of the resulting segmentation can be defined.

In practice, the purpose of nodule segmentation is often growth rate determination. It has been observed [1] that volume errors in automatic segmentation methods are not necessarily problematic for accurate growth rate estimation as long as they are consistent. The current method should be tested on more data to determine its reproducibility and accuracy in growth rate determination.
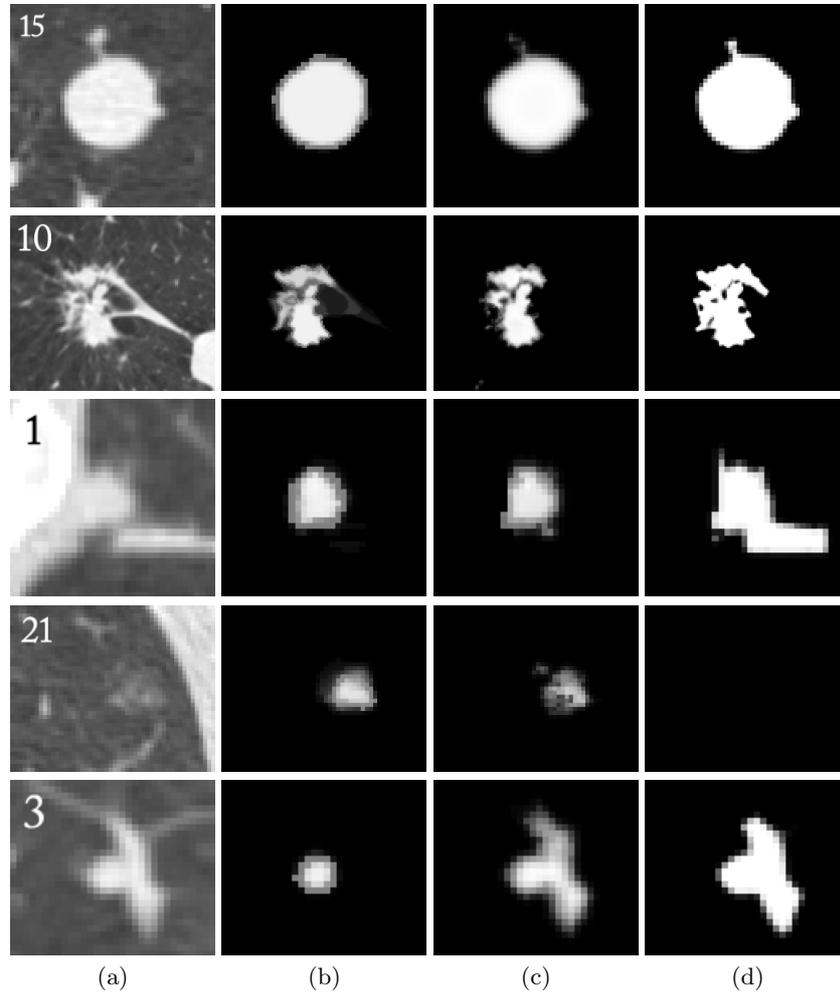
**Fig. 1.** Segmentation result for various cases. (a) CT data, axial or coronal sections zoomed in on the nodule with a lung window level, W/L = 1400/-600 HU. (b) Truth. (c) Results of the proposed method. (d) Result of conventional method [2].

## 6    Conclusion

The manual outlines in the LIDC data set make a strong case for using soft instead of hard segmentation schemes for pulmonary nodules. The proposed method is fully automated, provides soft segmentations and learns from examples, contrary to existing approaches that use hard-coded rules, typically require correct choices of internal parameters to achieve good results, and produce binary segmentations. Although the system is not extensively optimized and trained with only a small number of cases, the obtained results are very promising.

# References

1. Mullally, W., Betke, M., Wang, J.: Segmentation of nodules on chest computed tomography for growth assessment. Med. Phys. **31** (2004) 839–848
2. Kostis, W., Reeves, A., Yankelevitz, D., Henschke, C.: Three-dimensional segmentation and growth rate estimation of small pulmonary nodules in helical CT images. IEEE Trans. Med. Imag. **22**(10) (2003) 1259–1274
3. Fetita, C.I., Préteux, F., Beigelman-Aubry, C., Grenier, P.: 3D automated lung nodule segmentation in HRCT. In: MICCAI. Volume 2878 of LNCS. (2004) 626–634
4. Kuhnigk, J.M., Dicken, V., Bornemann, L., Wormanns, D., Krass, S., Peitgen, H.O.: Fast automated segmentation and reproducible volumetry of pulmonary metastases in CT-scans for therapy monitoring. In: MICCAI. Volume 3217 of LNCS. (2004) 933–941
5. Okada, K., Comaniciu, D., Krishnan, A.: Robust 3D segmentation of pulmonary nodules in multislice CT images. In: MICCAI. Volume 3217 of LNCS. (2004) 881–889
6. Okada, K., Comaniciu, D., Krishnan, A.: Robust anisotropic Gaussian fitting for volumetric characterization of pulmonary nodules in multislice CT. IEEE Trans. Med. Imag. **24** (2005) 409–423
7. Wiemker, R., Rogalla, P., Blaffert, T., Sifri, D., Hay, O., Shah, E., Truyen, R., Fleiter, T.: Aspects of computer-aided detection (CAD) and volumetry of pulmonary nodules using multislice CT. British Journal of Radiology, **78** (2005) 46–56
8. Sluimer, I., Schilham, A., Prokop, M., van Ginneken, B.: Computer analysis of computed tomography scans of the lung: a survey. IEEE Trans. Med. Imag. **25**(4) (2006) 385–405
9. Wormanns, D., Diederich, S.: Characterization of small pulmonary nodules by CT. Eur. Radiol. **14** (2004) 1380–1391
10. Armato, S.G., McLennan, G., McNitt-Gray, M.F., Meyer, C.R., Yankelevitz, D., Aberle, D.R., Henschke, C.I., Hoffman, E.A., Kazerooni, E.A., MacMahon, H., Reeves, A.P., Croft, B.Y., Clarke, L.P.: Lung Image Database Consortium: Developing a resource for the medical imaging research community. Radiology **232**(3) (2004) 739–748
11. Armato, S.G., Giger, M.L., MacMahon, H.: Automated detection of lung nodules in CT scans: Preliminary results. Med. Phys. **28**(8) (2001) 1552–1561