Multiscale Vessel-guided Airway Tree Segmentation

Pechin Lo[1], Jon Sporring[1], and Marleen de Bruijne[1,2]

[1] Image Group, Department of Computer Science, University of Copenhagen, Denmark, pechin@diku.dk,
[2] Biomedical Imaging Group Rotterdam, Departments of Radiology & Medical Informatics, Erasmus MC, Rotterdam, The Netherlands.

Abstract. This paper presents a method for airway tree segmentation that uses a combination of a trained airway appearance model, vessel and airway orientation information, and region growing. The method uses a voxel classification based appearance model, which involves the use of a classifier that is trained to differentiate between airway and non-airway voxels. Vessel and airway orientation information are used in the form of a vessel orientation similarity measure, which indicates how similar the orientation of the an airway candidate is to the orientation of the neighboring vessel. The method is evaluated within EXACT'09 on a diverse set of CT scans. Results show a favorable combination of a relatively large portion of the tree detected correctly with very few false positives.

1 Introduction

Most existing airway segmentation methods are based on region growing, with the assumption that the airway lumen has low intensity and is surrounded by higher intensity airway walls. The main problem with such an intensity based region growing algorithm is that the contrast between the airways and their surroundings is sometimes very low, due to noise or pathologies such as emphysema. Such low contrast regions often cause the region growing algorithm to leak into the surrounding lung tissue. Currently there are two approaches to address this problem: explosion control and the use of more advance image descriptors than intensity alone.

The idea of explosion control is to stop the segmentation in the low contrast regions where otherwise leakage would occur, while the segmentation continues in other regions. Strategies for explosion control generally involve heuristic rules based on geometrical properties of the regions labeled. Some examples of these geometrical properties are: volume of the regions segmented [1], radius of propagation front [2, 3], cross section area [4] and topology of thinned structure [5].

The second approach makes use of local image information to better differentiate between airways and their surroundings, for instance using pattern recognition techniques [6–8] or local tube fitting [9]. The method presented in this paper belongs to this second approach.

This paper presents an extension of our previous work [8], where we proposed to incorporate both a trained appearance model and the similarity between the orientation of an airway and its accompanying vessels. In this paper we use multi-scale Hessian eigen analysis instead of the fixed scale analysis as described in [8]. The method is evaluated within the EXACT'09 [10] airway extraction challenge on a database of 20 scans taken at different sites under a variety of different conditions. The results were manually evaluated by trained observers and compared to results submitted by other participants.

This paper is organized as follows: Section 2 explains how the airway appearance model is constructed using the training set. Section 3 presents the various steps involved in computing the vessel orientation similarity measure. The segmentation framework that combines both the airway appearance model and the vessel orientation similarity is presented in Section 4. Section 5 presents the results of the 20 cases in the EXACT'09 testing set. Finally, a discussion of the results and comparison with results submitted by other teams are presented in Section 6.

2 Classification based airway appearance model

2.1 Incomplete segmentation as a basis for training

One of the drawbacks of a classification-based appearance model is the need for training data. We have shown in [7] and [8] that incomplete but leakage free airway tree segmentations, which can be obtained relatively easily, can be used as a substitute for real ground truth segmentations as training data.

We obtain the needed manual segmentations using intensity based region growing, where both a seed point within the trachea and an intensity threshold are provided manually. The highest threshold possible without causing any leakage is selected for each training image individually. This typically results in an over-conservative segmentation that is incomplete, with many missing branches but has no leakage. As the 'background' regions directly surrounding such a conservative segmentation will always contain some airway voxels, an additional 'leaked segmentation', obtained using a slightly higher threshold, is used to take this into account. We use these leaked segmentations to prevent uncertain regions that may be either airway or background from being used in the training process. An example of a manual and a leaked segmentation is shown is Figure 1(a) and (b).

Before the extraction of training samples, we extract the lung fields, trachea and main bronchi using a thresholding and morphological smoothing based algorithm, as presented in [7, 8]. Training samples from two classes are extracted from the training data: the airway class and the non-airway class. The airway class consists of all voxels that are labeled in the manual segmentation, excluding the trachea and main bronchi. The non-airway class is limited to the area that is within the lung fields and close to the airways, which are obtained by first dilating the manual airway segmentation with a sphere of radius R_{dilate}. The

non-airway class then consists of the area within this dilated region that are not marked by the leaked segmentation.

To ensure approximately independent training samples, only a small percentage S_{sample} of the voxels belonging to the airway class are used for training. The same number of training samples is also extracted from the non-airway class. In order to prevent the large number of voxels in the larger airways from dominating the appearance model, we sample evenly along the distance from the main bronchi, measured through the segmented tree. This is done by grouping the voxels based on their distance from the main bronchi in bins with the width of each bin fixed to W, and randomly sample a total of $N = S_{sample}VW/D_{max}$ training samples from each bin, where V is the total number of airway class voxels in the manual segmentation and D_{max} is the maximum distance between a voxel in the manual segmentation to the main bronchi. To ensure that we do not sample a bin too densely, at most 50% of all voxels belonging to a bin will be included. The sampling starts at the bin furthest away from the main bronchi. If the required number of samples from a bin is larger than the number of voxels available in the bin, the remaining samples are extracted from the next available bin of shorter distance.

2.2 Airway probability

The training samples extracted are used to train a k nearest neighbor (KNN) classifier [11] to differentiate between voxels belonging to the airway and non-airway classes. An initial set of local image descriptors or features is computed from the training samples, which consists of spatial derivatives up to and including the second order, eigenvalues of the Hessian matrix (λ_1, λ_2 and λ_3, where $|\lambda_1| \geq |\lambda_2| \geq |\lambda_3|$), determinant and trace of the Hessian matrix, Frobenius norm of the Hessian matrix, and combinations of Hessian eigenvalues that measure tube, plate and blobness ($|\lambda_2/\lambda_1|$, $|\lambda_3/\lambda_1|$, $(|\lambda_1| - |\lambda_2|)/(|\lambda_1| + |\lambda_2|)$, $|\lambda_3|/\sqrt{|\lambda_1\lambda_2|}$). The partial derivatives of the image are computed at multiple scales by convolving the image with the partial derivatives of a Gaussian kernel [12], and each of the features is standardized to zero mean and unit variance.

Sequential forward feature selection [13] is used to find an optimal set of image descriptors that maximizes the area under the receiver operating characteristic (ROC) curve of the classifier. To this end, the training samples are randomly partitioned into two parts to compute the ROC curve: one third for training of the classifier and two thirds for validation. The final KNN classifier is trained using the optimal combination of features and all the training samples.

We can now estimate for each voxel in previously unseen images the posterior probability of the voxel belonging to the airway class, given a set of optimal features $\boldsymbol{x}$, using the following:

$$p(A|\boldsymbol{x}) = \frac{K_A(\boldsymbol{x})}{K} \tag{1}$$

where A is the airway class, $K_A(\boldsymbol{x})$ is the number of neighbors around $\boldsymbol{x}$ that belong to the airway class, obtained among the K nearest neighbors.

3 Obtaining vessel orientation similarity

The vessels are segmented from the lung fields, using a multi-scale Hessian eigen analysis approach. The scale for calculating the Hessian matrix is selected for each voxel independently using the scale normalized [14] Frobenius norm of the Hessian matrix:

$$\omega(\sigma_i) = \sigma_i^2 \sqrt{\lambda_1(\sigma_i)^2 + \lambda_2(\sigma_i)^2 + \lambda_3(\sigma_i)^2}$$

The local vessel scale, σ_v, is then obtained by selecting the smallest scale that corresponds to a local maximum of $\omega(\sigma_i)$ across scales. Using the Hessian eigenvalues at scale σ_v, the following criteria are used to evaluate whether a voxel belongs to a vessel or not:

Brightness:	$\lambda_1, \lambda_2 < 0$
Contrast:	$\omega \geq T_\omega$
Tubeness 1:	$(\lvert\lambda_1\rvert - \lvert\lambda_2\rvert)/(\lvert\lambda_1\rvert + \lvert\lambda_2\rvert) < T_1$
Tubeness 2:	$(\lvert\lambda_1\rvert - \lvert\lambda_3\rvert)/(\lvert\lambda_1\rvert + \lvert\lambda_3\rvert) > T_2$

A voxel is labeled as vessel when it fulfills all four criteria. The brightness criterion ensures that only voxels that are brighter than their surroundings will be selected, the contrast criterion reduces the effects of noise by ensuring a certain minimum contrast between the voxel and its surroundings, and finally the two tubeness criteria require vessels to locally resemble bright, solid cylinders. Segmentation using the vessel criteria often results in additional small, isolated regions due to noise. A connected component analysis using a 6-connected neighborhood scheme is employed to remove regions that are smaller than V_{min} voxels. Finally, the vessel centerlines are obtained using the 3D thinning algorithm presented in [15].

The vessel orientation at the centerline voxels is obtained as the eigenvector corresponding to λ_3 computed at the vessel scale σ_v. This measure is less sensitive to noise and inaccuracies in the vessel segmentation than the orientation obtained directly from the centerline itself. The orientation of an airway is extracted the same way as the orientation of a vessel, through multi-scale Hessian eigen analysis. The Hessian matrix is constructed using the airway probability image, generated using (1) where the airways resemble solid bright tube structures. Given θ as the angle between the local tube orientation at an airway candidate voxel and the orientation measured at the centerline of a vessel nearest to it, the vessel orientation similarity is defined as $s = |cos(\theta)|$. When the two orientations are similar then $s \simeq 1$, and when the orientations are perpendicular then $s \simeq 0$.

4 Segmentation framework

The airway tree segmentation is obtained using a 3D region growing algorithm, with a decision function that combines both the airway appearance model of Section 2 and the vessel orientation similarity of Section 3. An initial airway

segmentation described in Section 4.1 is used as seeds for the region growing algorithm. Figure 1(f) shows a block diagram of the proposed segmentation framework.

4.1 Initialization

The segmentation process is initialized with a coarse segmentation of the first four airway generations obtained using intensity based region growing. First, starting from the trachea and main bronchi as obtained in Section 2.1, all connecting voxels with intensity lower than a threshold T_{airway} are added. This is followed by a closing operation with a sphere mask of 3 voxels in radius. Finally, an algorithm [2] that is capable of tracking generations via bifurcation detection is applied to the smoothed segmentation, and only branches up to the fourth generation are retained. The threshold T_{airway} is dynamically determined by searching from -1000HU, with an increment of 5 HU, until the resulting initial segmentation, excluding trachea and main bronchi, is larger than 1000 voxels.

4.2 Airway segmentation

The initial segmentation obtained previously is used as seed points in a region growing process to extract the remainder of the airway tree, using the airway probability and vessel orientation similarity measures. The vessel similarity is used as a means to relax the requirements on airway probability in regions with a local tube orientation that is similar to the orientation of nearby vessels. We achieve this by using the following decision function to decide on whether to accept an airway candidate voxel $\boldsymbol{x}$ or not

$$D(p(A|\boldsymbol{x}), s) = \begin{cases} 1, & p(A|\boldsymbol{x}) \geq T_u \\ 1, & T_u > p(A|\boldsymbol{x}) \geq T_l \text{ and } s \geq T_s \\ 0, & \text{otherwise,} \end{cases} \tag{2}$$

where $p(A|\boldsymbol{x})$ is the airway probability computed from (1), s is the vessel orientation similarity of the candidate voxel, T_u, T_l and T_s are the upper probability threshold, lower probability threshold, and vessel similarity threshold respectively. The voxel $\boldsymbol{x}$ is labeled as an airway when $D(p(A|\boldsymbol{x}), s) = 1$.

4.3 Parameter settings

KNN classification was performed using the ANN library for approximate nearest neighbor searching [16]. A $K = 21$ was used, and the approximation error *eps* was set to zero to turn off the approximation part of the algorithm. For the extraction of training samples, as described in Section 2.1, the dilation radius R_{dilate} was set to 5 mm, sampling percentage S_{sample} was set to 0.05, and the bin width W was set to 3 voxels. A total of 7 scales, distributed exponentially between 0.5 mm and 3.5 mm, were used to compute the features, as well as for

the multi-scale Hessian eigen analsysis in Section 3. A contrast threshold T_ω of 100, and tubeness measure thresholds T_1 and T_2 of 0.5 were used for the vessel criteria. Detected vessel regions smaller than $V_{min} = 20$ voxels were discarded. All these parameter values are the same as those reported in [8].

The thresholds for the decision function (2) was hand tuned based on the training set, where the airway probability images were generated with KNN classifiers that were constructed in a leave-one-out manner. Our experiments with the training set showed that $T_u = 0.86$, $T_l = 0.62$ and $T_s = 0.8$ gives good results, where a significant amount of new branches not in the training data were found without any significant leakages.

5 Experiments and results

Among the 20 cases (CASE01-CASE20) in the training set from EXACT'09, manual segmentations for training were successfully obtained from all cases except one (CASE06), where leakage was observed even when using the lowest possible threshold. Therefore only 19 cases from the training set were used to train the KNN based appearance model.

The training process of our method took approximately 13 hours on a single CPU of an Intel Xeon X5355 processor (2.66 GHz), with the feature selection process occupying around 11 hours. The average computation time to segment an image in the test case was approximately 1 hour and 30 minutes. Most of the computation time was spent generating the airway probability image, which took an average of 50 minutes. The remaining computation time was mainly spent on generating the Gaussian blurred images and performing the multi-scale Hessian eigen analysis.

Table 1 shows the evaluation results of our method on the 20 EXACT'09 test images. Surface renderings of the best case and worst case according to the detected tree length are shown in Figure 1(c) and (d).

6 Discussions and conclusion

At the expense of a relatively long computation time and laborious training procedure, we obtain a favorable combination of a relatively large portion of the tree detected correctly with very few false positives. Table 1 shows that the proposed method is able to extract at least 50% of the total tree length for 70% of the cases, with a false positive rate of less than 1% for all cases. Although many branches were either extracted only partly or missed completely, it should be noted that no method was able to extract more than, on average, 77% of tree length or branches in the ground truth. Among the 15 methods that were compared in EXACT'09, 7 methods resulted in both a lower tree length and a higher false positive rate. Compared to the remaining 7 methods, our results stand out mainly by the small amount of leakage; in 11 cases there were no false positives at all, and in the remaining cases both leakage volume and false positive rate were small. Figure 1(c) shows few clear leakages even in the case

Table 1. Evaluation measures for the twenty cases in the test set.

	Branch count	Branch detected (%)	Tree length (cm)	Tree length detected (%)	Leakage count	Leakage volume (mm^3)	False positive rate (%)
CASE21	114	57.3	64.2	58.1	0	0.0	0.00
CASE22	276	71.3	227.1	68.7	1	0.2	<0.01
CASE23	186	65.5	137.9	53.0	1	15.4	0.12
CASE24	128	68.8	106.0	65.1	0	0.0	0.00
CASE25	152	65.0	116.9	46.4	0	0.0	0.00
CASE26	48	60.0	32.9	50.0	0	0.0	0.00
CASE27	49	48.5	36.9	45.6	0	0.0	0.00
CASE28	77	62.6	57.3	52.2	0	0.0	0.00
CASE29	117	63.6	81.3	58.9	1	10.0	0.12
CASE30	140	71.8	109.5	71.7	3	26.2	0.32
CASE31	159	74.3	117.8	67.1	3	39.1	0.31
CASE32	151	64.8	120.9	55.5	2	7.3	0.05
CASE33	108	64.3	81.3	55.2	0	0.0	0.00
CASE34	301	65.7	213.1	59.6	3	14.0	0.07
CASE35	136	39.5	95.2	30.8	0	0.0	0.00
CASE36	187	51.4	185.2	44.9	0	0.0	0.00
CASE37	57	30.8	46.4	26.1	0	0.0	0.00
CASE38	36	36.7	27.8	41.8	0	0.0	0.00
CASE39	253	48.7	195.1	47.7	6	37.0	0.30
CASE40	333	85.6	315.5	81.5	17	214.8	0.90
Mean	150.4	59.8	118.4	54.0	1.9	18.2	0.11
Std. dev.	85.2	13.6	75.4	13.4	3.9	48.0	0.22
Min	36	30.8	27.8	26.1	0	0.0	0.00
1st quartile	77	48.7	57.3	45.6	0	0.0	0.00
Median	138	63.9	107.7	54.1	0	0.0	0.00
3rd quartile	253	71.3	195.1	67.1	3	26.2	0.30
Max	333	85.6	315.5	81.5	17	214.8	0.90

with the largest number of detected leaks (CASE40). Our method is also the only method with an average false positive rate below 1% (0.11%) that is still able to achieve an average detected tree length of higher than 50% (54%), as shown in Figure 1(g). Among all methods with an average false positive rate below 5%, the method achieving the highest tree length, which is a semi-automated method, detected only 5% more of the total tree length at a false positive rate of 1.19%.

The presented method was originally developed for segmenting the airway tree in more standardized, low-dose cancer screening scans obtained at a single site [17, 8]. While the main approach and parameter settings used for the current work are the same as in [8], the implementation differs from this previous work in three ways. Firstly, the Hessian eigen analysis to determine vessel and airway orientations, performed at a single resolution level in [8], was replaced by a

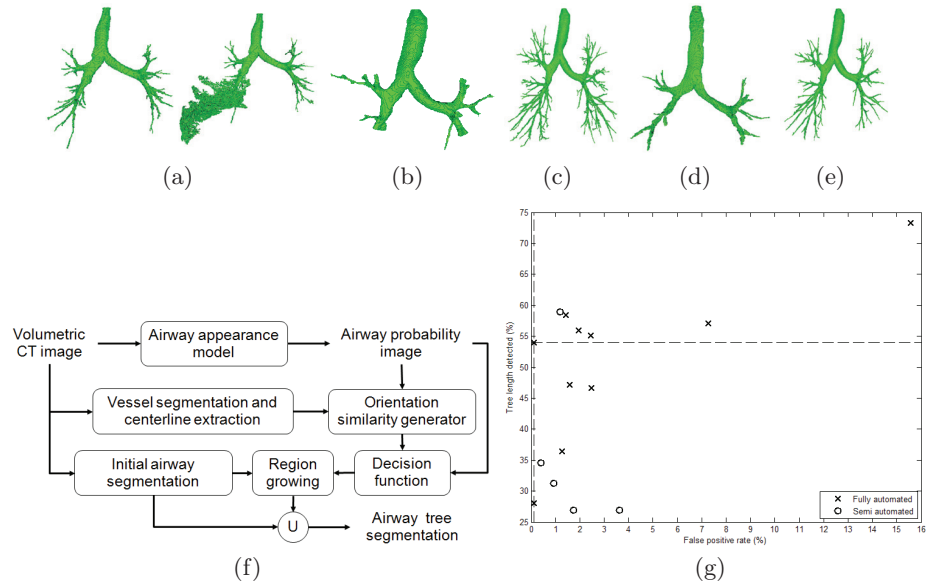

Fig. 1. Surface renderings of (a) a manual (left) and leaked (right) segmentation used for training, (b) initial segmentation, test results with (c) largest (CASE40) and (d) smallest (CASE37) percentage of tree detected. (e) Surface renderings of segmentation from CASE40 obtained using screening study scans from [8] as training. (f) Block diagram of the segmentation framework. (g) A scatter plot of average tree length detected versus average false positive rate of all participating teams in EXACT'09, with the proposed method at the intersection of the dashed lines.

multi-resolution analysis. This modification had already been developed for the screening study and improved results considerably for that data.

To cope with the much more diverse data of EXACT'09, the method was further modified in two ways from our experiments on the training images. The first modification is that a more complete segmentation is used to initialize the classification-based region growing, instead of using only the trachea and main bronchi as reported in [8]. The reason for this is because we found that otherwise the segmentation was sometimes already terminated within the first four generations in noisy images. The second modification is that the thresholds for the decision function in this work were manually tuned based on the training data, while those in our earlier work were tuned automatically using an automatic leakage detection algorithm similar to [2] and [3]. In the diverse and sometimes very noisy CT scans from EXACT'09, this rule-based, automatic leakage detection algorithm turned out to be unreliable. The criterion used in selecting the thresholds was that no obvious leakage should be present in the results on the training set, which is probably one of the reasons why our results are on the conservative side compared to other methods. A different set of thresholds may lead to longer tree lengths at the expense of an increase in false positives.

In the EXACT'09 study, a training set was provided with carefully selected images that were of similar quality as the images in the test set. If the training data does not match the test data, results may deteriorate. To illustrate this, segmentation of CASE40, obtained using a Siemens Sensation 16 scanner and very sharp convolution kernel B70s, was repeated with the same setup and same parameter settings but using the training data of [8], which consisted only of scans obtained using a Philips Mx8000 IDT 16 scanner with softer kernel D, resulting in less noisy images. The result is shown in Figure 1(e). The method trained with different data has still little or no leakage, but finds fewer branches. In this case, this was mainly because the appearance model is not capable of handling the noise and often misclassifies bright noise voxels in the airway lumen as non-airway. Although the method should be trained on similar data for optimal results, the good results on the diverse set of EXACT'09 data indicate that application of this method is not limited to studies in which acquisition conditions can be standardized.

Note that, although the proposed method requires training data, in this work only very low quality segmentations were available for this purpose. The segmentations used for training consisted of on average 93 branches and had a total length of 99 cm. Application of the trained models on the test set of similar images resulted already in much more complete segmentations, with 150 branches and a total length of 118 cm. Clearly, our method can achieve better results than the training segmentations, but the lack of training examples from small branches does limit the ability of the system to extract higher generation airways. With the availability of a set of high quality segmentations for training, such as the ground truth resulting from EXACT'09, we expect to obtain even better results in the future.

In conclusion, an airway segmentation method that uses a voxel classification based appearance model and the similarity between the orientation of an airway and its neighboring vessel is presented. Compared to the results from other algorithms submitted to EXACT'09, our method is especially effective in avoiding leakage, while still being able to extract a fair amount of airway branches.

Acknowledgments. This work is partly funded by the Danish Council for Strategic Research (NABIIT), the Netherlands Organization for Scientific Research (NWO), and AstraZeneca, Lund, Sweden.

References

1. Kiraly, A.P., Higgins, W.E., Hoffman, E.A., McLennan, G., Reinhardt, J.M.: 3D human airway segmentation for virtual bronchoscopy. In: SPIE Medical Imaging 2002: Physiology and Function from Multidimensional Images. Volume 4683. (April 2002) 16–29
2. Schlathölter, T., Lorenz, C., Carlsen, I.C., Renisch, S., Deschamps, T.: Simultaneous segmentation and tree reconstruction of the airways for virtual bronchoscopy. Volume 4684., SPIE (2002) 103–113

3. van Ginneken, B., Baggerman, W., van Rikxoort, E.: Robust segmentation and anatomical labeling of the airway tree from thoracic CT scans. In: Medical Image Computing and Computer-Assisted Intervention. Volume 5241 of Lecture Notes in Computer Science. (2008) 219–226
4. Kitasaka, T., Mori, K., Suenaga, Y., Hasegawa, J., Toriwaki, J.: A method for segmenting bronchial trees from 3D chest X-ray CT images. In: MICCAI (2). (2003) 603–610
5. Tschirren, J., Hoffman, E., McLennan, G., Sonka, M.: Intrathoracic airway trees: segmentation and airway morphology analysis from low-dose CT scans. Medical Imaging, IEEE Transactions on **24**(12) (Dec. 2005) 1529–1539
6. Ochs, R.A., Goldin, J.G., Abtin, F., Kim, H.J., Brown, K., Batra, P., Roback, D., McNitt-Gray, M.F., Brown, M.S.: Automated classification of lung bronchovascular anatomy in CT using AdaBoost. Medical Image Analysis **11**(3) (June 2007) 315–324
7. Lo, P., de Bruijne, M.: Voxel classification based airway tree segmentation. In: Medical Imaging 2008: Image Processing. Volume 6914., SPIE (2008) 69141K
8. Lo, P., Sporring, J., Ashraf, H., Pedersen, J., de Bruijne., M.: Vessel-guided airway segmentation based on voxel classification. In Brown, M., de Bruijne, M., van Ginneken, B., Kiraly, A., Kuhnigk, J., Lorenz, C., Mori, K., Reinhardt, J., eds.: Proc. of First International Workshop on Pulmonary Image Analysis. (2008)
9. Graham, M.W., Gibbs, J.D., Higgins, W.E.: Robust system for human airway-tree segmentation. In: Medical Imaging 2008: Image Processing. Volume 6914., SPIE (2008) 69141J
10. Lo, P., van Ginneken, B., Reinhardt, J., de Bruijne, M.: Extraction of airways from ct (exact09). In: Second International Workshop on Pulmonary Image Analysis. (2009)
11. Duda, R.O., Hart, P.E., Stork, D.G.: 4.4. In: Pattern Classification. 2 edn. Wiley-Interscience (2001) 174 – 177
12. Weickert, J., Ishikawa, S., Imiya, A.: On the history of Gaussian scale-space axiomatics. In Sporring, J., Nielsen, M., Florack, L., Johansen, P., eds.: Gaussian Scale-Space Theory. Kluwer Academic Publishers, Dordrecht, The Netherlands (1997) 45–59
13. Pudil, P., Novovičová, J., Kittler, J.: Floating search methods in feature selection. Pattern Recogn. Lett. **15**(11) (1994) 1119–1125
14. Lindeberg, T.: Feature detection with automatic scale selection. Int. J. Comput. Vision **30**(2) (1998) 79–116
15. Wang, T., Basu, A.: A note on 'A fully parallel 3D thinning algorithm and its applications'. Pattern Recognition Letters **28**(4) (March 2007) 501–506
16. Arya, S., Mount, D.M., Netanyahu, N.S., Silverman, R., Wu, A.Y.: An optimal algorithm for approximate nearest neighbor searching fixed dimensions. J. ACM **45**(6) (1998) 891–923
17. Pedersen, J., Ashraf, H., Dirksen, A., Bach, K., Hansen, H., Toennesen, P., Thorsen, H., Brodersen, J., Skov, B., Døssing, M., Mortensen, J., Richter, K., Clementsen, P., Seersholm, N.: The danish randomized lung cancer CT screening trial - overall design and results of the prevalence round. Journal of Thoracic Oncology (April 2009)