COMPUTER-AIDED DETECTION OF PULMONARY NODULES FROM CT SCANS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Sergey Victorovich Fotin

May 2011

COMPUTER-AIDED DETECTION OF PULMONARY NODULES FROM CT SCANS

Sergey Victorovich Fotin, Ph.D.

Cornell University 2011

Lung cancer is the leading cause of cancer related death in the world. Computed tomography (CT), which can provide detailed images of lung structure, makes it possible to detect lung cancer in its early stage. Regular clinical practice involves visual inspection of hundreds cross-sectional slices of a patient's CT scan for small pulmonary nodules that can manifest early lung cancer. However, radiologists routinely miss nodules due to fatigue and the error-prone nature of the work, which may ultimately lead to incorrect diagnostic decisions. It has been shown that detection performance can be improved significantly by employing a computer algorithm for pulmonary nodule identification.

This dissertation is devoted to the topic of computer-aided detection (CAD) of pulmonary nodules from chest CT scans. The thesis includes several subtopics: system architecture, optimization and validation of the detection system. Among the major contributions to the topic are: design and development of a multiscale Laplacian of Gaussian-based candidate generation system, high specificity standard moments-based pulmonary vessel bifurcation filter, non-solid nodule detection system, and a new detection system validation procedure that compensates for size measurement error and provides a more meaningful performance assessment for CAD systems. In addition, a large size-enriched dataset for CAD system evaluation was created to become a valuable resource for future research.

BIOGRAPHICAL SKETCH

Since 2005, Sergei Fotin has been working with the Vision and Image Analysis Research Group at Cornell University. Prior coming to Cornell, he studied Software and Computer Engineering at Ural State Technical University, Yekaterinburg, Russia.

To my family

ACKNOWLEDGEMENTS

I would like to thank my academic adviser, Anthony Reeves, for his major support and encouragement during this research work. Most importantly, I am grateful for his teaching, academic guidance and invaluable advice on writing, making clear presentations and giving good talks. I feel very fortunate to have a teacher whose opinion I genuinely respect.

I am thankful to Peter Doerschuk and Noah Snavely, for their insights on some of the aspects of this research and being patient members of my special committee; Claudia Henschke, David Yankelevitz and Matthew Cham, for their clinical outlook and great help with collecting the data; Andinet Enquobahrie, for the assistance during the initial phase of this project; Tatiyana Apanasovich, for help with the statistical data analysis methods. I greatly value the instruction I have received from outstanding Cornell teachers: Rich Caruana, Martin Burtscher, Sheila Hemami, Dan Huttenlocher and Ramin Zabih.

I have been happy to have amazing Rhodes 357 officemates. I am thankful to Alberto Biancardi, Andy Browder, Eric Heinz, Artit Jirapatnakul, Brad Keller, Jaesung Lee, Jeremiah Wala, Sander Becx for being good colleagues and friends at the same time.

Special thanks go to my friends Ilya and Filipp, for introducing me to Cornell; Leonid, for the best time; Victor, for being himself; and Lisa, for helping me balance the most important things in life.

Finally, this research would not be possible without support from National Institutes of Health (grant R33CA101110) and the Flight Attendants' Medical Research Institute.

Detect nodules, we must!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION: PROBLEM OF COMPUTER-AIDED DETECTION

Lung cancer is currently considered the most deadly form of cancer, accounting for more deaths than breast cancer, prostate cancer, and colon cancer combined. It is estimated that 157,300 people died from lung cancer in the United Stated in 2010[1]. Early detection of lung cancer by radiologist through the use of medical imaging techniques, such as chest X-ray and conventional tomography, has long been thought to improve prognosis by allowing malignant nodules to be caught in earlier stages, when treatment is considered most effective. The advent of computed tomography (CT) by Hounsfield [1973] has allowed for improved diagnosis of lung cancer by providing even more detailed images of the lung structure [Muhm et al., 1977; Schaner et al., 1978; Meziane et al., 1988], which in turn increases the likelihood of a malignant lesion being detected. In addition, as CT technology and image resolution improves, lesions are able to be resolved in greater detail and the minimum detectable lesion size is reduced [Fischbach et al., 2003].

However, as CT image resolution improves, the increased amount of information presented to a radiologist turns visual inspection of the entire CT volume into a time consuming task [Rubin, 2000; Fleishon et al., 2006]. This has led to issues where both throughput and accuracy are compromised due to both increasing read-time required per scan, and associated reader fatigue. To address these issues, computer-aided detection (CAD) systems have been proposed that would, ideally, reduce the amount of information presented to radiologists by focusing a radiologist's attention on only those portions of an image that are

[1]American Cancer Society. Cancer Facts and Figures, 2010. `http://www.cancer.org/`

likely to contain malignant or suspicious lesions.

1.1 Definition and objective of CAD

Typically, the radiological evaluation of a medical image consists of two main stages: (a) detection of suspicious regions from an image volume and (b) evaluation of the identified suspicious regions. Within the context of computerized image analysis, these stages correspond to computer-aided detection (CADe, as often referred in the literature) and computer-aided diagnosis (CADx). Even though these procedures are very similar in nature and share some components, they solve different problems and it is important to discriminate between the two. This thesis focuses on the first stage, computer-aided detection, or the procedure that automatically identifies pulmonary nodules from a whole-lung CT volume without further diagnosis. For the remainder of this thesis, the abbreviation ”CAD” will be used to denote ”computer-aided detection.”

The main task of a lung CAD system is finding abnormal structures inside otherwise normal CT images of human lungs. The challenge is to identify small nodules that, in the midst of a very large number of pulmonary structures, are very difficult to see, as illustrated in Figure 1.1; i.e. a needle in a haystack problem.

1.2 Pulmonary nodule types

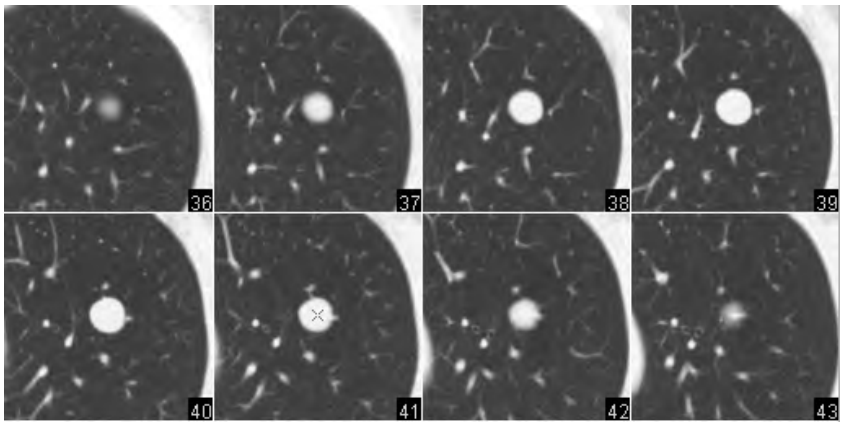
The task of nodule detection is complicated by the nonuniform presentations of pulmonary nodules, which may have different consistencies and geometrical

Figure 1.1: A slice of a chest CT image showing multiple similar looking pulmonary structures with only one being a true nodule (marked by thick arrow).

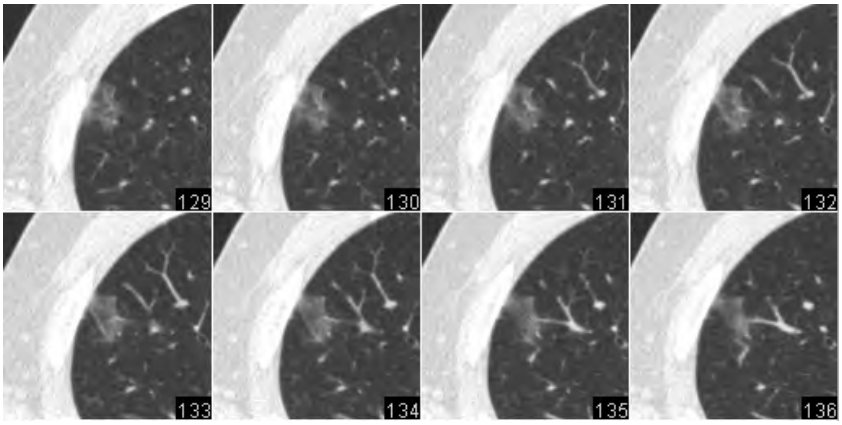
morphologies.

From pathology and appearance, pulmonary nodules can be classified into two categories: solid and nonsolid. Solid pulmonary nodules, such as shown in Figure 1.2(a), are a relatively common finding. They are opaque, mostly spherical structures that can be located anywhere in the lung. Solid nodules are represented by a cluster of voxels that have an intensity range similar to other structures that can be seen on chest CT: pulmonary blood vessels, thick airway walls, and chest wall.

From the geometrical shape perspective, solid nodules can be further catego-



(a)



(b)

Figure 1.2: Example of (a) solid and (b) nonsolid nodules.

rized by the degree and character of the attachment to a neighboring structure as shown in Figure 1.3[1]. Such variety makes the detection task difficult, necessitating sophisticated computer algorithms. For example, nodules adjacent to major blood vessels are of particular difficulty, as the detection algorithm must be able to distinguish this complex geometry from regular blood vessels or vessel bifurcation points. Similarly, if a nodule is attached to the lung pleura, it

[1]Here montage images on the right side are supplemented with light shaded three-dimensional visualizations of example nodules on the right. To obtain each visualization, (a) the 3D region surrounding the nodule of interest was manually marked and clipped from the image; (b) the intensity threshold of -400 HU (Hounsfield Units) was applied to this region; (c) a marching cube algorithm was used to construct a 3D mesh; (d) this surface was rendered and captured to the image.

needs to be distinguished from a natural surface irregularity.

Nonsolid pulmonary nodules, also referred to as ground-glass opacities (GGO), are another class of findings. As can be seen in Figure 1.2(b), they are less dense than solid nodules, and do not completely obscure the lung parenchyma. Their image intensity range is much lower than solid nodules, but is comparable to image artifacts caused by heart and respiratory motion. Both the consistency and shape of nonsolid lesions makes them quite distinct volumetric structures within the lungs. However, because of their transparent appearance, the region of a nonsolid nodule may contain dense pulmonary vessels and airways that may complicate their detection. Finally, nonsolid lesions have been shown to have a lower incidence rate and possibly higher chance of malignancy than solid nodules [Henschke et al., 2002].

Part-solid nodules are another subclass of pulmonary nodules that usually have a central solid component surrounded by a nonsolid component. From the automated detection point of view, they can be treated as either solid or nonsolid depending on the amount of their solid part regardless of their clinical classification.

1.3 Minimum nodule size criterion

Target nodule size range is usually determined from a clinical protocol. In this research, the protocols of three lung screening studies were considered: IELCAP [Henschke et al., 1999], NELSON [Xu et al., 2006] and ITALUNG [Lopes Pegna et al., 2009]. These studies were designed to test the effectiveness of screening for early lung cancer using low-dose CT. High risk

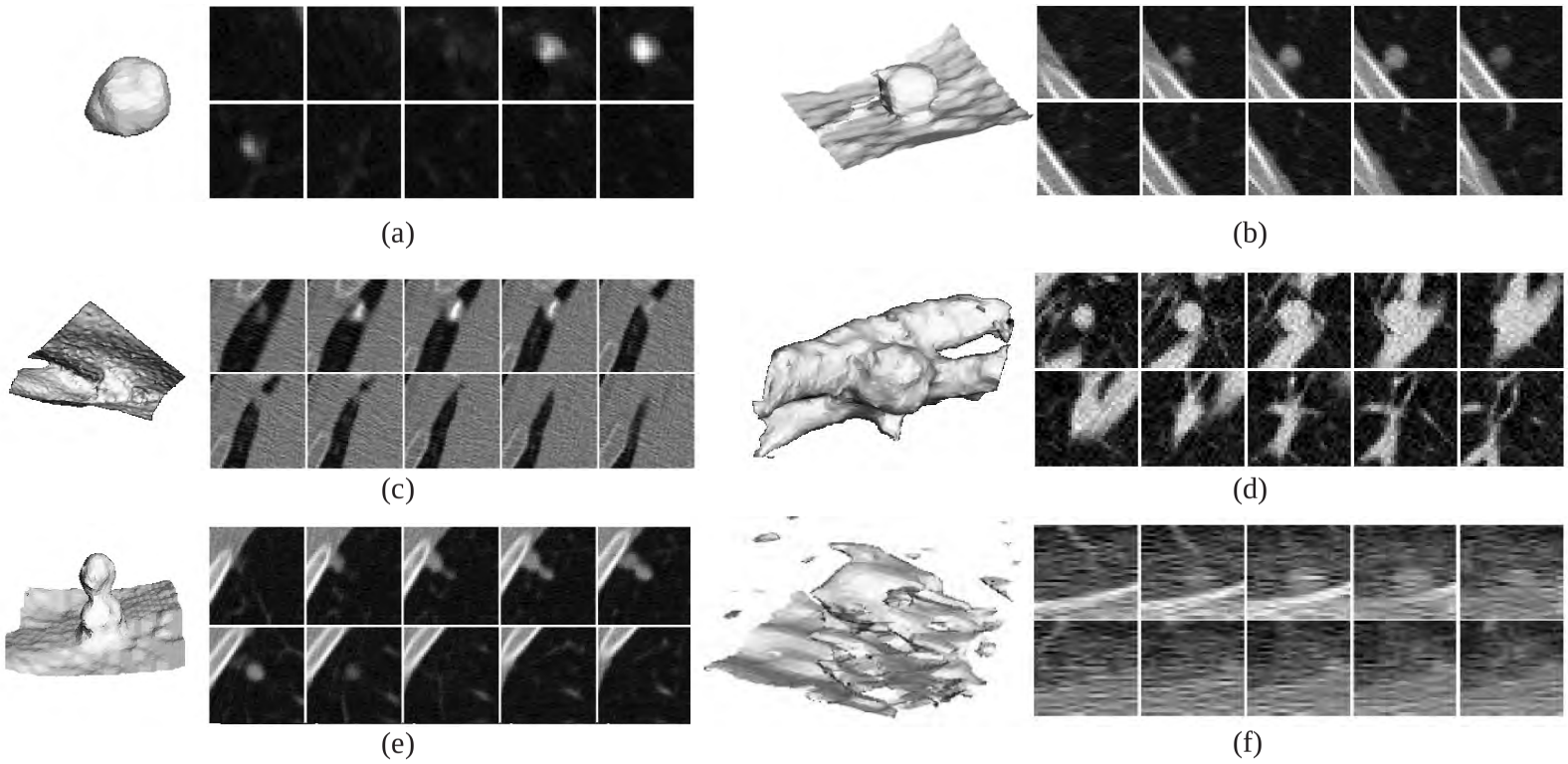


Figure 1.3: Nodule presentations on CT image: (a) regular isolated nodule; (b) nodule attached to pleural surface; (c) costophrenic nodule (placed between two lung walls in diaphragm region); (d) nodule with attachments to pulmonary vessel and airway; (e) nodule of unusual shape; (f) nodule surrounded by noise artifacts.

Table 1.1: Selection of target nodule minimum size from clinical protocol specification (minimum nodule size is specified in millimeters).

Study	Solid nodules	Nonsolid nodules
IELCAP	5	8
NELSON	4.5	8
ITALUNG	5	10

subjects were enrolled in each study and the baseline CT scan was acquired. The result of an examination was considered non-negative if a pulmonary nodule that met a minumum size criterion was identified. The minimum nodule size criteria for these studies are shown in Table 1.1.

From the perspective of a CAD system, it is also reasonable to set up the minimum cut-off sizes to be in agreement with a clinical protocol. In practice it means that the CAD system algorithm has to reject nodule candidates that do not meet the minimum size criterion.

1.4 Typical lung CAD scheme

Most algorithms for nodule detection from CT scans reported in the literature conform to the common design shown in Figure 1.4. This is a standard pattern recognition scheme that usually consists of four major steps: image preprocessing, nodule candidate generation, feature extraction and classification of candidates.

In the first step, the CT image is prepared for the generation of nodule can-

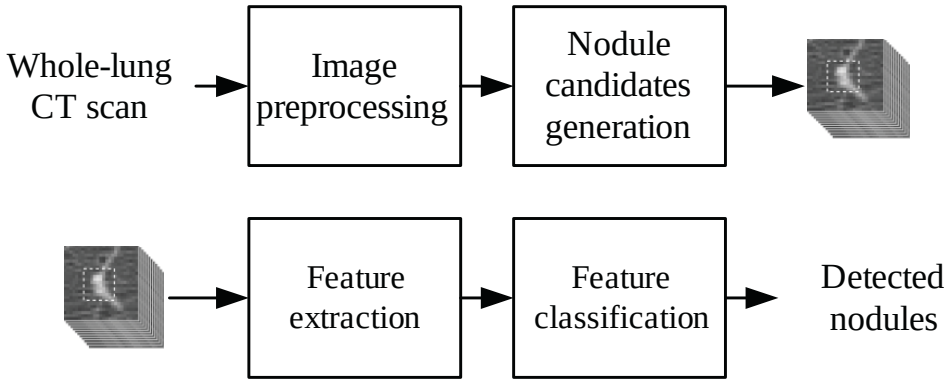


Figure 1.4: A flowchart of a typical lung nodule detection system.

didates. This preparation may include multiple image processing techniques, such as resampling, denoising etc. Often the three-dimensional search space, the region of the image where pulmonary nodules may be located, is identified. The second step is nodule candidate generation, where the system inspects the whole CT scan and identifies all objects (candidates) that morphologically resemble lung nodules. In addition to geometrical location, the nodule size may be estimated and recorded for each candidate. The third step involves extraction of image features for each identified candidate. In this final step, false positives (FPs) are eliminated using feature classification, and the set of detected nodules is presented to an operator.

1.5 Performance evaluation

The main performance measures of a CAD system are (a) sensitivity, the fraction of correctly identified nodules and (b) false positive rate (FP rate), the average number of findings that are not nodules per case (or per patient)[1]. Both mea-

[1]In earlier work on nodule detection from thick slice CT, the false positive rate was reported per slice (or section)

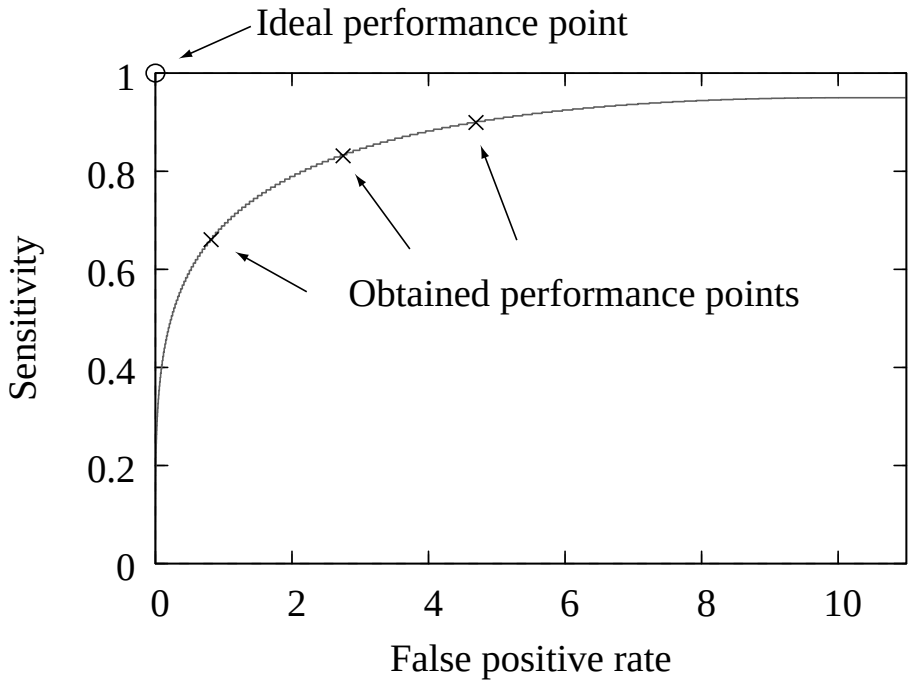


Figure 1.5: A typical free-response receiver operating characteristic (FROC) curve displaying the trade-off between detection sensitivity and false positive rate.

sures are often used together in performance evaluation, and are derived from a comparison of an expert decision to the decision of the detection algorithm. The pairs of sensitivity and false positive rate can be plotted on a two-dimensional graph called a free-response receiver operating characteristic (FROC) [Miller, 1969]. The FROC is the standard and widely accepted tool for performance measurement in many detection domains. A FROC curve displays the trade-off between the sensitivity and false positive rate, as shown in Figure 1.5. The aim of an automated detection system is to simultaneously maximize sensitivity and minimize the false positive rate, that is, produce a curve passing as close to the left upper corner of the FROC space as possible.

The typical validation procedure for a CAD system requires the algorithm to

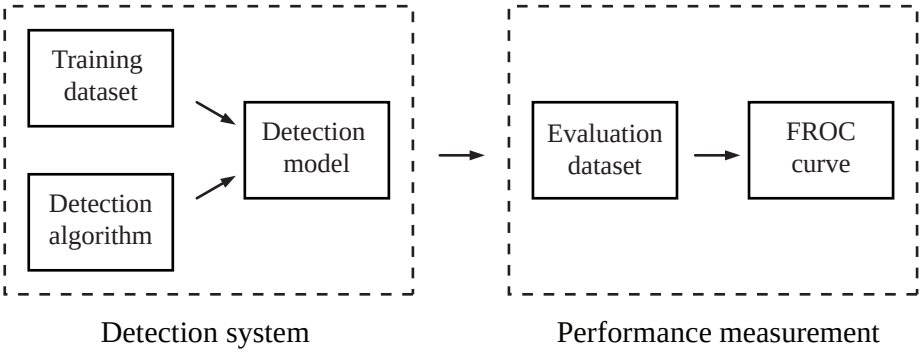


Figure 1.6: An illustration of the performance evaluation procedure.

be optimized (trained) and evaluated (tested) on separate sets of CT scans. This is summarized in Figure 1.6.

Optimization of a CAD algorithm involves the construction of multiple detection models that depend both on used detection algorithm parameters and on the training set. Only the best detection models with the best candidate classification accuracy must be selected, i.e. the models with maximal sensitivity and minimal false positive rate. These models are typically found using statistical learning and classification methods. Within this paradigm, selection of the best models' parameters is translated to selection of the best classifier and its parameters. The convex hull of obtained performance pairs (sensitivity, false positive rate) makes up the training FROC curve, which consists of a set of optimal operating points. Indeed, each operating point corresponds to certain parameters of a CAD system or/and classifier. Once a set of operating points is obtained, corresponding detection models are applied to the evaluation set and the resulting performance is calculated. The result of applying the best models' parameters to the evaluation set is reflected in the final FROC curve.

There are two main factors that need to be considered when comparing dif-

ferent systems using FROC curves: (a) target nodules types and sizes; (b) properties of evaluation datasets. As CAD systems are usually developed to target certain nodule types and sizes, a reported FROC curve cannot be used to predict the behavior of the system on a different nodule target range. Similarly, if the system was developed on a dataset that has certain properties (CT scan resolution, radiation dose, patient selection strategy etc.), one is likely to obtain a different FROC curve from a dataset with other properties. This complicates the direct comparison of various detection systems.

1.6 Previous work

Previous studies have shown that a standalone CAD system may result in a nodule detection sensitivity comparable to manual detection by an experienced radiologist [Wormanns et al., 2004; Yuan et al., 2006], which falls in the range between 30% and 70%. When a CAD system and a radiologist interpret CT scans independently, the overall detection sensitivity is always greater than one produced by the radiologist alone [Awai et al., 2004; Marten et al., 2005; Das et al., 2006; Sahiner et al., 2007a; Beyer et al., 2007]. Promising results on datasets of substantial size are shown by both nodule detection systems developed within the academic environment [Arimura et al., 2004; Farag et al., 2004; Armato et al., 2005; Sahiner et al., 2007a; Pu et al., 2008; Li et al., 2008; Murphy et al., 2009], and within industry [Marten et al., 2005; Wormanns et al., 2004; Yuan et al., 2006]. Surveys on the recent works within the area can be found in the papers by Sluimer et al. [2006]; Li [2007]; Chan et al. [2008]; Van Ginneken [2008]. A brief summary of selected CAD systems (where provided), is shown in Table 1.2, and corresponding FROC operating points are plotted in Figure 1.7. Although this

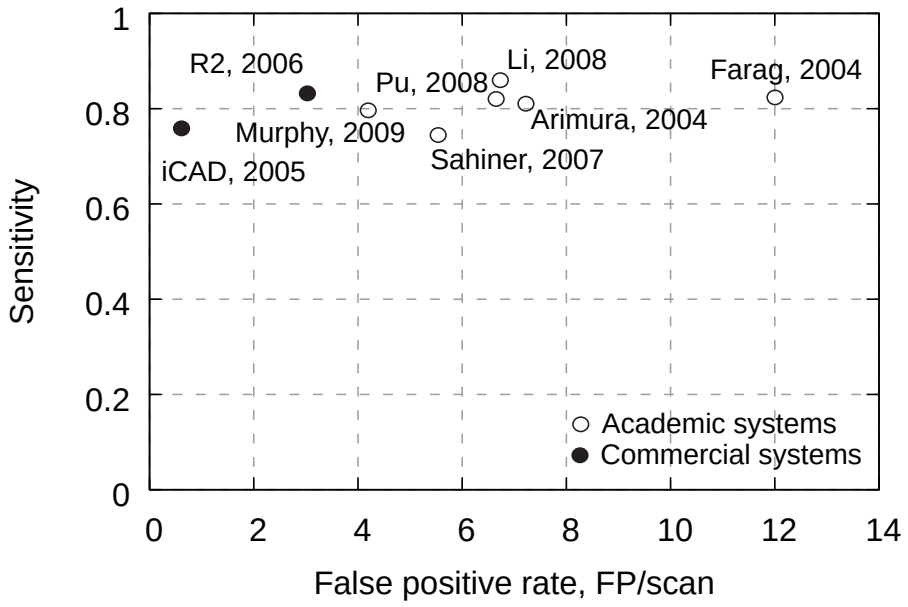


Figure 1.7: Performance of various reported detection performance on a FROC plot.

graph cannot be considered a fair comparison between the CAD systems, it does allow for a high-level overview of the current state of the field.

1.6.1 Literature review

The graph in Figure 1.7 shows that sensitivities of the CAD systems fall between 0.7 and 0.9, while commercial systems tend to have a lower false positive rate. Even though commercial systems have been found to be more competitive, the details of the algorithms are not fully disclosed to the general public. For this reason, only a limited high-level review for some of the CAD systems that were developed in academia will be provided. The following systems were selected from a pool of hundreds of papers based on sizes of the datasets used for algorithm evaluation.

Table 1.2: Detection systems reported in the literature: evaluation datasets and obtained performance.

First author	Year	Number of cases	Slice thickness, mm	Number of nodules	Size range, mm	Nodule types	Sensitivity, %	False postive rate, FP/scan
Kanazawa	1998	450	10	230	≥ 4	n/a	90	n/a
Arimura	2004	106	10	131	3 - 26	n/a	81	7.3
Farag	2004	200	n/a	130	n/a	n/a	82	12
Wormanns*	2004	n/a	1.25	457	1 - 18 (mean = 3.9)	all	55	33 total
Armato	2005	393	10	470	3 - 30 (mode $\approx$ 6)	all	70	1.6 FP/slice
Marten*	2005	20	various	135	1 - 29.6 (median = 4.4)	all	76	0.55
Yuan*	2006	150	1.25	337	≥ 4	all	83	3.18
Sahiner	2007	85	n/a	122	3 - 88 (median = 5.7)	n/a	77	5.5
Pu	2008	52	2.5	184	3 - 29	all	82	6.5
Li	2008	117	1 - 2	153	4 - 28 (mode $\approx$ 6)	all	86	6.6
Murphy	2009	813	0.7	3981	≥ 4	all	80	4.2

* detection systems developed in industry

The CAD system developed by Arimura et al. [2004] achieved a sensitivity of 83% with a 5.8 false positive rate on a low-dose CT scans dataset containing 109 nodules. The lung search space was found by applying intensity thresholding on a CT scan followed by a morphological closing operation for vessel removal. Prior to candidate generation, the original scan was filtered by a nodule enhancement filter that was developed and had been found to work well on chest radiographs. Candidates were identified using a multiple-level thresholding technique. False positive candidates were eliminated by applying ”multi- massive trained” artificial neural networks and linear discriminant analysis classifiers to the extracted image features.

The system presented by Farag et al. [2004] segments the lungs and major pulmonary vessels as the first step of the algorithm. Matching with templates of different shapes was used to identify the nodule candidates. For this reason the authors designed four different nodule templates with shape parameters that could be altered. Nodule identification was done by encoding the location and size of possible nodule candidates in a ”chromosome.” By altering individual bits, hypothetical candidate locations were matched with templates using normalized cross-correlation. Classification of the candidates was done with the help of a Bayesian classifier trained on three image features. The system achieved a sensitivity of 82.3% with a false positive rate of 12 on a database consisting of 130 nodules.

The algorithm described by Armato et al. [2005] was evaluated on 393 cases containing 470 pulmonary nodules. The sensitivity was 70%, while the false positive rate was 1.6 per slice. Lung area was segmented by thresholding with a value determined from the gray-value histogram. Nodules were identified by

thresholding with gradually decreasing values similar to Arimura et al. [2004]. The classification of nodule candidates was based on a combination of rule-based and linear discriminant classifiers applied to a set of two- and three-dimensional features.

The CAD system designed by Li et al. [2008] was evaluated on 117 CT scans containing 153 nodules and achieved a sensitivity of 86% with 6.6 false positives per scan. The scheme of the algorithm follows the standard pattern recognition scheme containing lung segmentation, candidate generation, feature extraction and discrimination of candidates. The uniqueness of this approach is in using selective filters for nodule enhancement and suppression of normal anatomic structures such as blood vessels. These filters are multiscale and are able to characterize local image structures as spheres, vessels or lung walls. Such discrimination is possible due to computing local image curvature using second-order image derivatives.

Pu et al. [2008] obtained a sensitivity of 81.5% with a false positive rate of 6.5 per case on a dataset of 52 CT scans containing 184 nodules. After constructing the lung mask as a search space for nodule candidates, the binary image of a CT scan was obtained using an empirically selected threshold. The nodule candidate generation approach used by the authors is quite simple — a signed Euclidean distance field was calculated and local maxima with corresponding sizes were recorded as nodule candidates. Finally, shape-based analysis was performed to reduce false positive findings.

The work presented by Murphy et al. [2009] is distinct from the other studies as the authors used 813 CT scans from a NELSON lung cancer screening trial [Xu et al., 2006] for evaluation of their system. On data containing 3981 nodules,

the sensitivity achieved by the CAD system was 80% and the false positive rate was 4.2 per case. As usual, the algorithm starts with lung segmentation and proceeds with candidate generation in the identified volume. Candidates are identified by computing shape indicies and curvedness for each image location. Voxels having these values above a certain threshold are then clustered together and adjusted to form nodule candidates. The set of false positive candidates was further reduced by means of classifying image features with a k-nearest neighbor classifier.

As we can see, all mentioned systems have the same architectures reliant on lung field segmentation, candidate generation, feature extraction and classification. The most significant differences between the algorithms are in the candidate generation scheme and the pools of used image features.

The majority of the detection algorithms are optimized for solid nodules, which are more common, do not perform well on nonsolid nodules. The low incidence of nonsolid lesions makes it difficult to establish a large documented image database for CAD development, which may account for the limited attention that they have received in the literature.

Several different detection techniques that specifically target nonsolid nodules have been published recently. Kim et al. [2005] suggested a two-dimensional analysis method where Gaussian fitting and texture features were calculated over overlapping square patches covering the lung field. An artificial neural network was used for the final classification. The method was tested on 29 nonsolid regions and achieved a sensitivity of 90% and a false positive (FP) rate of 0.89 per section. Ye et al. [2007] used a multiscale dot enhancement filter followed by adaptive thresholding and a rule-based classifier based on distance

maps, motion tracking and local features. They evaluated 50 scans containing 52 nodules and obtained a 92.3% sensitivity and 12.7 FPs/scan. Yang et al. [2007] designed a method for the reduction of false positives for nonsolid nodule detection using volumetric shape descriptors computed over 3D volumes of interest. Evaluation was done on a set of 216 cases containing a mixture of 243 solid and 81 nonsolid lesions and resulted in an 81% sensitivity and 4.3 FPs/scan. Bastawrous et al. [2007] employed another false positive reduction technique based on 2D cross-sectional analysis and achieved a 96.3% sensitivity and 0.06 FPs/slice rate for the set of 27 nonsolid nodules.

In spite of the difficulties in comparing the results reported by various research groups, the detection performance numbers tend to improve over the years. In the year 2000, thick-slice image acquisitions were dominant and many researchers treated CT volume as a set of about 30 2D slices, and reported false positive per slice. As the image quality improved and CT scans reached submillimeter isotropic resolution, the majority of research teams switched to a per scan false positive count. As evidence for the overall progress in detection performance, the false positive rate in 2000 expressed per slice, and in 2010 expressed per CT volume, is of the same order.

1.6.2 Outstanding challenges

As the performance of CAD systems improves over time, the challenges remain the same. In this section, both algorithmic and organizational difficulties faced by CAD system developers are discussed.

None of the reviewed authors report the sensitivity of a 1.0, which means

that some of the nodules identified by humans are always missed by CAD. This is one of the limitations of modern lung CAD systems that makes it unacceptable to use them as a sole reader in a clinical setting. The main obstacle in a development of the system with ideal sensitivity is sub-optimal candidate generation. As it will be made further evident in Chapter 5, the challenge of constructing perfect candidate generators remains open.

Very few algorithms have been published on nonsolid nodule detection. One possible reason for this is that nonsolid nodules are a relatively infrequent finding. Often there are few nonsolid nodules in the available evaluation sets, so the CAD systems are designed either to ignore them or to detect them using the already existing solid nodule detection approach. From a performance point of view, the contribution of nonsolid nodules to overall sensitivity is small and it may make sense simply to disregard them. However, for a high-sensitivity detection system, performance on the nonsolid nodules is critical.

Another algorithmic challenge reported by many researchers is large nodules. Visible shape variation increases with nodule size and often it is difficult to distinguish large nodules from normal lung anatomy. Unless explicit anatomical knowledge is accessible to the algorithm, detecting large nodules will remain challenging. On the one hand, some may argue that large nodules should not be a target for lung CAD because these nodules are easily identified by a human observer. On the other hand, when a CAD system is a sole reader, it must identify all nodules according to predefined specifications.

False positive rates for high sensitivity levels are still high. Vessels and vessel bifurcations are universally acknowledged as being the major source of false positive findings by a lung CAD system. Local vessel tree anatomy often re-

sembles the anatomy of a vascularized nodule. This makes it difficult for both human and computer observers to distinguish between the two.

One of the major organizational issues is the absence of common agreement among the scientific community about what constitutes an evaluation dataset for a lung CAD system. A majority of the researchers considers any structure identified by a radiologist (or a consensus of radiologists) as a nodule to be detected. Due to human subjectivity, opinions of radiologists may differ, thus introducing an uncertainty for system design and evaluation. Moreover, size and type (solid vs. nonsolid, benign vs. malignant) constraints are often imposed on target nodules by a clinical protocol. The procedures of size measurement and nodule type assignment introduce another uncertainty for CAD system developers. In an ideal scenario, a CAD system should be able to redefine its output according to the well-defined target nodule specifications.

Certification of a CAD system for use in clinical practice remains a challenge for commercial CAD developers in the United States. Limited diagnostic use of lung CAD systems and the absence of medical insurance coverage for automated interpretation of chest CT exams slow down the deployment of commercial CAD systems. The cost of Food and Drug Administration (FDA) approval and associated studies prevent many potential CAD vendors from entering the market. As of 2010, the FDA has approved two lung CADs (developed by R2 Technologies/Hologic and Siemens) for use as a second reader[1]. Other major vendors sell CAD systems that may assist in detecting nodules without claims approved by the FDA.

[1]Use of CAD system as a second reader involves these steps: (a) radiologist reads the CT image; (b) CAD system independently reads the CT image; (c) radiologist reviews CAD findings and makes the final report.

1.7 Research contributions

In this section, we consider the most important contributions and additions that have been made to address some of the challenges of the lung CAD mentioned before.

The documented dataset of CT images has been significantly increased in size. The main reason for this was to enlarge the size of training and evaluation sets and enrich the data with cases containing nodules of larger sizes. This allowed us to improve the generalization of our classification algorithm and increase performance on nodules of clinically relevant sizes. The details of the database development will be discussed in Chapter 2 of this thesis.

The CAD system architecture, which is discussed in Chapters 3 and 4, was extended with a module responsible for nonsolid nodule detection. The main motivation was to increase overall detection sensitivity of existing CAD systems that performed poorly on nonsolid nodules.

One way of increasing overall sensitivity is to use an effective candidate generation scheme. A multiscale normalized Laplacian of Gaussian filter, presented in Chapter 5, is capable of generating both solid and nonsolid nodule candidates with close to 100% sensitivity.

In an attempt to reduce the number of false positives, an algorithm specifically targeting the most common cause of false positives — pulmonary vessel bifurcations — was developed. A moment-based vessel bifurcation filter resulted in reduction of false positives of up to 69% in solid nodule detection. This filter will be described in Chapter 6.

Moreover, a method for the quantification of nodule size measurement error, which affects the calculation of both the sensitivity and false positive rate, was proposed. The developed scheme allows one to obtain a more meaningful performance assessment for CAD systems. The details of this scheme are given in Chapter 7.

Finally, several experiments were conducted to evaluate the performance of the CAD system on datasets with different properties. This evaluation is discussed in Chapter 8.

CHAPTER 2

DATASETS FOR TRAINING AND PERFORMANCE EVALUATION

Ideally, every CAD algorithm will go through development and validation stages in which its detection sensitivity and false positive rate are assessed. Properties of the evaluation datasets would dramatically affect design and validation of a lung CAD system.

Size of the evaluation dataset is probably the most important characteristic of the evaluation dataset that affects the robustness of the validation. For example, if the dataset consists of few CT scans or few nodules, there is a chance that particular detection system will be over-optimized for this data and would perform poorly on new, unseen data. To improve the generalization of a CAD algorithm, one may consider using evaluation sets of larger size containing a larger number of nodules.

Nodule size range is another important characteristic of a dataset that may affect the performance of a CAD system. If the nodules of target size range determined by a clinical protocol are not well-represented in the dataset, there is very little chance that the algorithm will produce any meaningful results. For example, if the development and evaluation of a CAD system is done on symptomatic cases, the system will likely be biased towards detecting larger, late-stage cancers. As a result, it will not be suitable for assessing CT scans in a screening setting where smaller nodules are more frequent.

A CAD system evaluated on a dataset with certain image acquisition parameters may also have flaws when exposed to datasets with different parameters. For example, early nodule detection systems were developed using thick-slice

data using mostly 2D image processing and analysis methods. Nowadays, CT scans approach isotropic resolution and submillimeter slice thickness — these make 3D methods more efficient and old CAD systems obsolete. The opposite effect is also true: a system developed using high resolution isotropic data will likely perform poorly on thick-slice data.

2.1 Sources of data

For the abovementioned reasons, we created an evaluation dataset of significant size that was further enriched with nodules of clinically relevant sizes. The sources of our data were diversified to minimize the algorithm ”overfitting” to particular image acquisition parameters. As the dataset has been accumulated gradually over the years, certain sections of this thesis only explored a subset of the entire CT scan dataset. The details of used data are clarified for each experiment individually. Parameters of the datasets used in development of our CAD system are summarized in Table 2.1.

Baseline system development and evaluation was done with respect to a size-enriched dataset of 706 whole-lung low-dose CT scans 1.25 mm thick from the Weill Cornell Medical Center database (CLDD 1.25). This dataset comes from the International Early Lung Cancer Action Program (I-ELCAP) [Henschke et al., 1999], which studies early lung cancer screening. The patients enrolled in the study were asymptomatic with indications for participation including age and smoking history. All scans in this dataset were reviewed by at least two thoracic radiologists and every found nodule was documented.

The original CLDD 1.25 dataset used by our research group consisted of 250

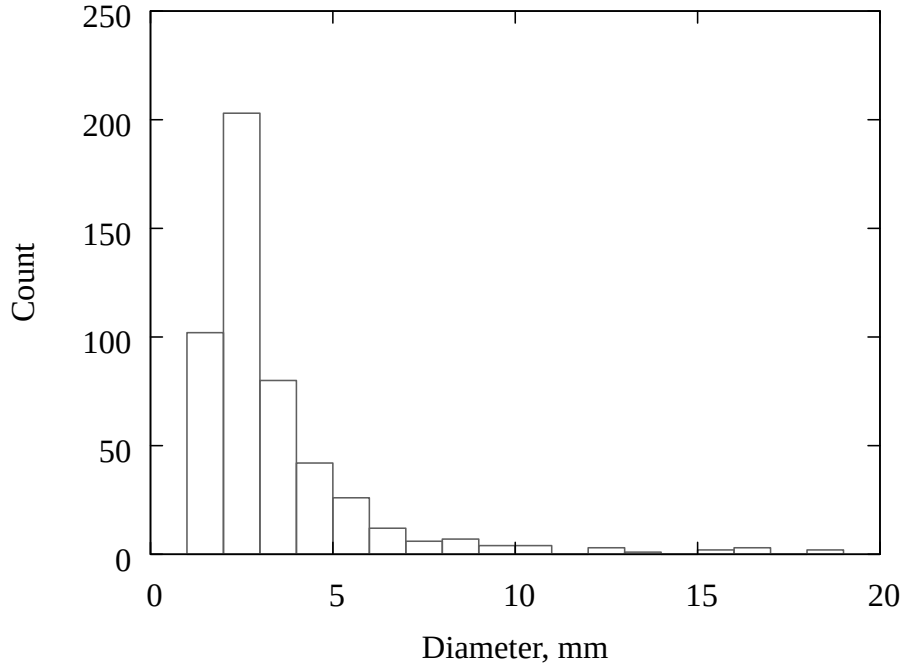
Table 2.1: Datasets for training and evaluation: description and image parameters

Dataset	CLDD 1.25	CLDD 2.5	LIDC
Source	Weill Cornell Medical Center		Publicly available dataset
Number of CT scans	706	250	350
Slice thickness, mm	1.25	2.5	0.625 - 0.3 (mean = 1.57, median = 1.80)
In-plane resolution, mm	0.49 - 0.89	0.51 - 0.80	0.48 - 0.76
CT scanner models	GE Lightspeed VCT/Pro16/Ultra		Various: GE, Siemens, Toshiba, Phillips
Beam	120 kVp, 40 - 80 mA	140 kVp, 40 - 50 mA	120 - 140 kVp, 40 - 422 mA
Convolution kernel	BONE		Various

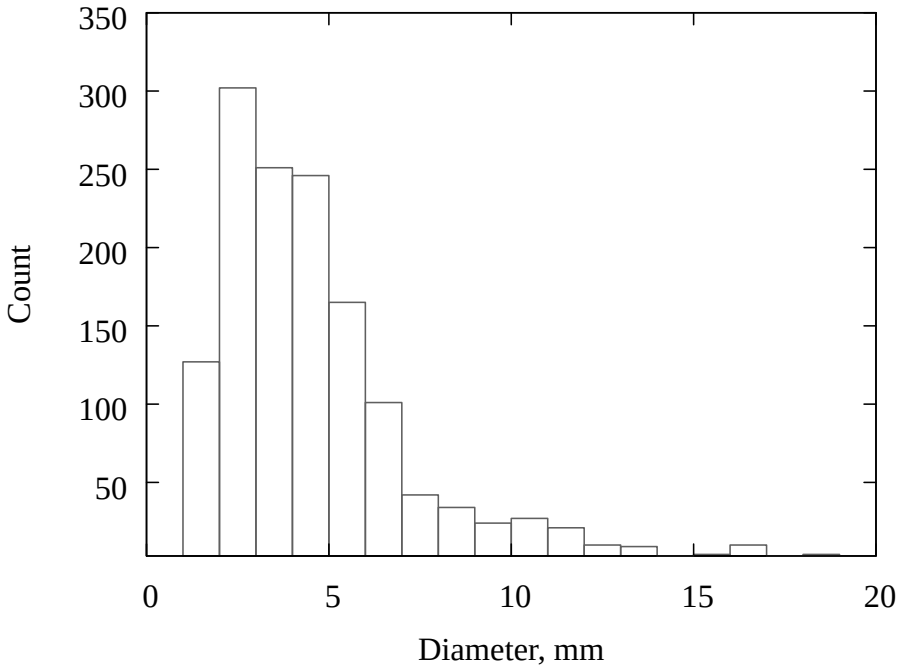
sequential cases [Enquobahrie et al., 2007], where tiny lesions with diameter smaller than 4 mm were dominant. To enhance the dataset, it was enriched with 456 new cases containing at least either one solid nodule greater than 4 mm or one nonsolid nodule greater than 6 mm in diameter as estimated by radiologists. These lower bound criteria for enrichment were selected to incorporate more nodules of the target size range. After the database was enriched, the fraction of nodules greater than 10 mm increased from 3% to 6% and the fraction of nodules greater than 4 mm increased from 23% to 43%. Distributions of the nodules in original and enriched datasets are shown in Figure 2.1. Such improvement of the evaluation sets allowed us to extend and test the capabilities of our detection system on the nodules that are more clinically important while providing better accuracy in performance evaluation.

Another dataset (CLDD 2.5) that was used for performance evaluation was from the same source, but only contained cases with the a slice thickness of 2.5 mm.

In addition to the CLDD datasets, a subset of 350 documented CT scans from the Lung Image Database Consortium (LIDC) [Armato III et al., 2004] dataset was used. Unlike the CLDD datasets which contained scans with slice thickness of 1.25 mm or 2.5 mm only, the LIDC set had scans of various slice thicknesses. Moreover, the LIDC CT scans were not obtained using a specific protocol, and were produced by CT machines from different vendors. As a result, images had much higher variation in radiation dose. When evaluating the detection performance on the LIDC set, the candidate nodule was considered as true if at least two radiologists marked it during the CT reading. Some of the cases in released data had incomplete or inconsistent markings, so 350 cases were used

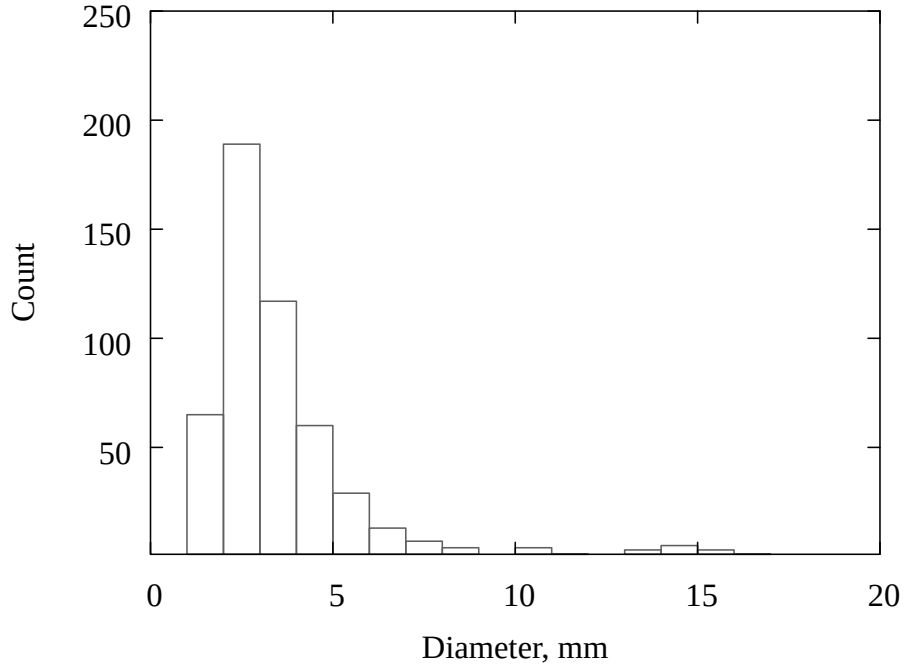


(a)

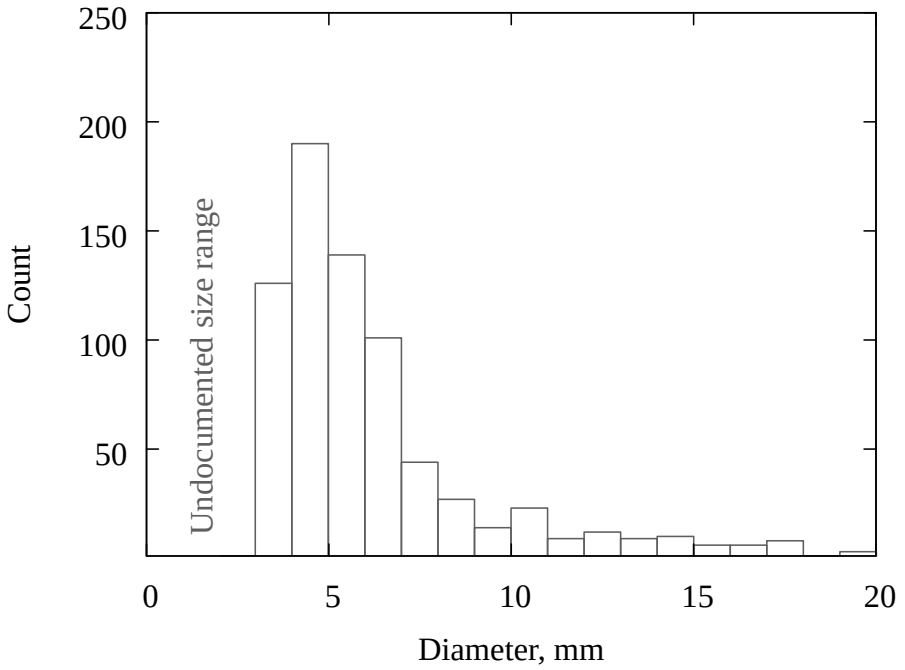


(b)

Figure 2.1: Distribution of nodules in the original (a) and enriched (b) CLDD 1.25 datasets.



(a)



(b)

Figure 2.2: Distribution of nodules in the CLDD 2.5 (a) and the LIDC (b) datasets.

for evaluation. The distribution of nodule sizes for the CLDD 2.5 and LIDC datasets are shown in Figures 2.2(a) and 2.2(b).

CHAPTER 3

DETECTION SYSTEM ARCHITECTURE

Because of the stated differences between solid and nonsolid nodules, detecting each of these classes can be done using specifically designed algorithms. In fact, detection of solid nodules can be reduced to the problem of finding and discriminating solid structures of different shapes, while detection of nonsolid nodules is equivalent to identifying spatial structures that have a certain intensity range. Because of these differences, a nodule detection CAD system system must have two separate processing subsystems responsible for each of the classes.

The flowchart of the detection system is shown in Figure 3.1. It has two parallel processing subsystems that are designed to detect solid and nonsolid nodules. Both subsystems are organized in a traditional way: nodule candidates are identified from the segmented lungs and then false positive findings are eliminated by classifying candidates using extracted features. The only architectural difference is that the nonsolid nodule detection subsystem requires an additional image preprocessing step before nodule candidate generation. The purpose of this step is to eliminate the interference from high intensity structures and is discussed in Chapter 5 of this thesis. After the candidates are generated, the set of features unique to each class is computed and candidates are classified. Remaining positive candidates from both subsystems are combined into the set of detected nodules, which is then presented to the operator. Depending on desired operating point, the trade off between sensitivity and the false positive rate can be altered. The system also allows the operator to see

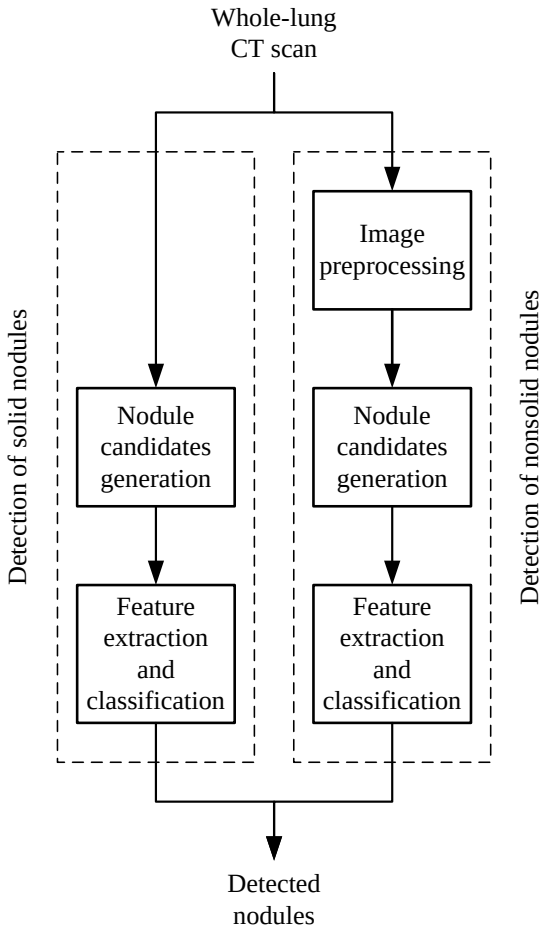


Figure 3.1: A scheme for detection of pulmonary nodules, containing two separate units for detecting solid and nonsolid nodules in parallel.

nodules within a specified size range.

3.1 Nodule candidates generation

The candidate generator is a key component of a nodule detection system. First, it must provide close to 100% initial sensitivity for all solid nodules, independently of their size and geometrical morphology. Second, in order to facilitate subsequent image feature extraction, it should provide a precise geometrical

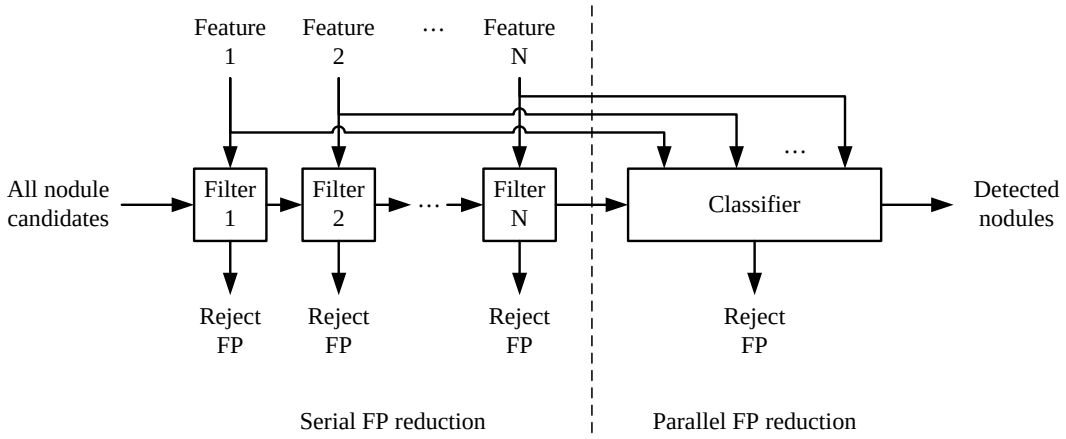


Figure 3.2: A scheme for false positive reduction and classification of solid pulmonary nodules: set of serial filters are followed by a parallel filter based on a multivariate classifier.

description of each candidate, including accurate centroid location and size estimates. Finally, it must be computationally efficient. All these requirements are satisfied by a candidate generator based on normalized Laplacian of Gaussian (LoG) filtering. The methods of generating both solid and nonsolid nodules are essentially the same (with the exception of the preprocessing step) and described and explored in Chapter 5 of this thesis.

3.2 Sequential false positive reduction

With a very large number of nodule candidates, feature extraction and classification must be computationally efficient. For this reason, a two-stage process for false positive elimination, shown in Figure 3.2, is proposed. The nodule candidates from the generator are first processed serially through a cascade of filters. When a very small portion of candidates from the original set remains, a parallel classification is used.

The sequential false positive reduction system has a similar design to the detection cascade published by Viola and Jones [2001] in the domain of face detection. Each of the filters used in serial false positive reduction takes in the set of remaining nodule candidates, and each one extracts a single feature from candidate definition and image information. Features are designed such that higher values favor true nodules. The candidates that have a relatively low feature value are then eliminated by thresholding. If we denote the number of candidates at the input of i-th serial filter as n_{i-1}, and the set of nodule candidates as $\{c_j^{i-1}\}_{j=1,n_{i-1}}$, then the output of the i-th filter will consist of n_i ($n_i \leq n_{i-1}$) candidates:

$$\{c_k^i\}_{k=1,n_i} = \{c_j^{i-1} : f_i(c_j^{i-1}) > T_i\}_{j=1,n_{i-1}}. \tag{3.1}$$

Before the application of the threshold T_i, the feature value $f_i(c_j^{i-1})$ must be computed. Since feature extraction and filtering is done in a sequential manner, the computational efficiency is maximized by computing fast features first. This way, the number of candidates for the processing is optimally reduced in terms of time for each step. This cascaded organization of filters saves a significant amount of time compared with traditional feature extraction and classification techniques.

Once the sequential false positive reduction steps are completed, only n_N candidates remain from an initial set of cardinality n_0, initially provided by the generator. After the final stage of sequential filtering, the remaining candidates already have features $f_1 \dots f_N$ computed, and they are used as an input to the parallel filter. This filter is a multivariate classifier that takes in all the features and produces the result of multivariate regression by assigning each candidate c_j^N a measure $r(c_j^N)$ approximately proportional to the probability of it being a

true nodule. Depending on the required detection sensitivity and false positive rate, an operator may set a threshold τ on the output of the classifier and obtain the set $\{C\}$ of detected nodules:

$$\{C\} = \{c_j^N : r(c_j^N) > \tau\}_{j=1,n_N}. \tag{3.2}$$

3.3 Features for solid nodule candidates

During feature generation, we assume that CT scans are 3D images and design the features to be purely three-dimensional and invariant to change in image resolution. A list of candidates from the candidate generator, containing centroid location and approximate size, is given to the first feature computation algorithm. Once the first feature is computed, the list is updated with the corresponding value and passed further to the next feature computation stage. In general, after the computation of a feature, the list of candidates is reduced in size by cutting off the candidates that have extreme values of the computed feature. This is described in Section 4.2.

The first feature used for classification is the normalized LoG filter response, which is explicitly used by the candidate generator to filter out parenchyma noise and lung wall irregularities. Another feature is the contrast, which is derived from application of the LoG filter of appropriate size to the candidate centroid. This feature is equal to the ratio of the weighted response in the positive part to the negative part of the LoG volumes of support.

The total attachment feature Enquobahrie [2007] indicates the degree of the candidate attachment to other pulmonary structures. When the value of the

attachment is too high, it means that the candidate is likely a lung surface irregularity or a major blood vessel coming out of the mediastinal region of the lungs. The secondary attachment feature is proportional to the second largest attachment of the candidate. This feature helps to discriminate between some attached nodules and pulmonary vessels, as true nodules rarely have more than one large attachment point.

The distances to the segmented lung boundary and to the airway tree, combined with estimated nodule candidate size feature, serve as indicators of whether the candidate is isolated or attached to either lung wall or airway vessel, respectively. Given this knowledge, attachment features help to refine the degree of parenchymal isolation for a given candidate.

Shape analysis of the candidate shape is then performed. As the very first step, the second-derivate features are computed as described in Li et al. [2003] to distinguish between round, elongated, and flat candidates. Finally, a moment-based feature gives higher-order information about the candidate that allows discrimination of vascularized nodules from pulmonary vessel bifurcations. Development and optimization of the moment-based feature is discussed in Chapter 6.

3.4 Features for nonsolid nodule candidates

As will be evident from Chapter 5, the altered nonsolid candidate generator produces high response to some solid nodules as well. In order to exclude solid candidates from the candidate list, the distribution of intensities within the volume for each candidate is analyzed from the original scan. For each candidate

c_j, the amount of nonsolid component is calculated as the fraction of voxels with intensities in the range -900 to -300 HU to the total volume of the candidate, and denoted as T_{frac}:

$$T_{frac} = \frac{|\{\overline{x} : \overline{x} \in c_j, I(\overline{x}) \in \text{(-900 HU, -300 HU)}\}|}{|\{\overline{x} : \overline{x} \in c_j\}|}, \tag{3.3}$$

where $I(\overline{x})$ is the image intensity of a voxel at location $\overline{x}$. These intensity values were selected as an initial approximation for intensity bounds for nonsolid tissue. Subsequently, candidates with a fraction below a certain threshold were rejected while only nonsolid candidates remained.

The normalized LoG filter response T_{resp} from the candidate generator and nonsolid fraction T_{frac} thresholds were used as classification features. As we can see, unlike the solid nodule detection subsystem, the nonsolid one requires computation of only two relatively simple features. As a result, in the current configuration, nonsolid nodule detection can be completed much faster than solid nodule detection.

3.5 Parallel classification

Selection of the classification algorithm is an important consideration in the design of a detection system. It is desirable to have a classifier that will provide regression output for each of the candidates. This way a simple threshold would determine the position of the operating point which can be altered to construct a FROC curve. The disadvantage of using a conventional classifier is that it provides just a binary decision and results in a single operating point in the FROC space.

For the task of classification, a soft margin support vector machine (SVM) [Vapnik, 2000] with linear kernel was used. SVMs are a family of feature classification algorithms that construct a hyperplane that will optimally separate positive and negative examples. An optimal hyperplane divides the feature space so that the distance (margin) to the closest negative and positive data samples is maximized. The larger this margin, the better the generalization of the classifier. Soft margin SVMs have been shown to result in low generalization error, even if trained on a small number of samples. Soft margin parameter c was optimized during SVM training. The distance to the resultant hyperplane was treated as a regression function.

In addition to the SVM, a distance weighted nearest neighbor (dwNN) classifier with a negative power weighing function [Atkeson et al., 1997] was used in one of the experiments. Prior to classification, the feature space was normalized in each dimension by scaling each feature by its variance. The classes of neighbor candidates were assigned the weights inversely proportional to the negative power p of the distance to the candidate being classified. The weighted sum was used as a final regression function.

During training there was no goal to obtain maximum possible absolute performance, but rather to see if a particular configuration of the system has an effect on its final performance. For this reason, extensive classifier optimization techniques such as feature selection and normalization, boosting, bootstrap sampling etc., were not used.

Given the set of labeled training samples, both classifiers produced a decision model. In the case of the SVM, the model describes the hyperplane dividing the opposite samples, while in the case of the dwNN classifier, it is a continuous

field with the potential at each point equal to the weighted sum of surrounding points' classes. Given the model description and new point in the test set, either the distance to the hyperplane or its potential is found.

In training, the only altered parameter was soft margin c and negative weighting power p for the SVM and dwNN classifiers, respectively. For obtaining the performance of the SVM, the optimal hyperplane was constructed using all candidates from the training set. For obtaining performance of the dwNN classifier on the training set, full leave-one-out cross-validation (LOOCV) was performed. The parameter that resulted in the best FROC curve for the training set was then used to produce the final classification model.

CHAPTER 4

OPTIMIZATION AND TRAINING

The system architecture described in the previous chapter has many parameters. By selecting these paramenters, both detection performance and computation time can be adjusted. Filter thresholds dertermine the speed and the bounds of detection sensitivity. Classification model parameters and regression function threshold determine the shape of the FROC curve and placement of operating point, respectively. In the following section, the general procedure for selecting these parameters is described in detail.

4.1 Nodules – candidates assignment

For the evaluation of detection performance, the candidates produced by the candidate generator should be assigned a classification of $True$ or $False$ based on the definition of nodules in the ground truth. For each nodule documented in the ground truth, a closest candidate was found and assigned a class $True$, if its center was located within the geometrical extent of the nodule. Otherwise, the candidate was marked as $False$. Since the performance of an automated pulmonary nodule detection systems is typically qualified with respect to some minimum size D_{th} of nodule to be detected, smaller candidates and nodules should be discarded from performance evaluation. However, due to size measurement error, nodule sizes may be either underestimated (UE) or overestimated (OE) by an automated system with respect to the ground truth measurements. To overcome this issue, detection performance is computed with the measurement error tolerance, which is described in Chapter 7 of this thesis. Nodules-candidates assignment and construction of the FROC curve are

identical for both solid and nonsolid detection modules.

4.2 Selecting filter thresholds

Optimization of the system for the best possible detection performance is an important design step. Since the solid and nonsolid nodule detection algorithms were designed independently, the optimization of each subsystem is done on the nodules of corresponding type within the training subset. The algorithms for feature generation were given (with the exception of the moment-based feature) and did not require additional fine-tuning; the solid nodule detection subsystem would require just an optimization of the thresholds for each of the sequential filters and training of the final classifier.

Each sequential filter within the solid detection subsystem has a threshold parameter T_i that separates input candidates into positives and negatives. Selection of these thresholds is crucial since each of them may eliminate candidates from any further consideration. Low, conservative values of this threshold would result in a higher number of candidates passed through the filter, thus providing higher sensitivity. However, such low thresholds would lead to ineffective use of filters: increased time for subsequent feature computation and, possibly, a higher final false positive rate. Finding a threshold that results in the best sensitivity-false positive rate trade-off for each of the sequential filters is an important part of the system optimization.

Each of the sequential filters' thresholds T_i is determined subjectively by looking at the distribution of feature values for true candidates and determining which true candidates have unreasonably low values of the feature. These true

candidates become outliers and the threshold is selected:

$$T_i = min\{f_i(c_j^{i-1}) : c_j^{i-1} = True,\ c_j^{i-1} \neq Outlier\}_{j=1,n_{i-1}} \tag{4.1}$$

Selection of the thresholds for each of the features is explained further in Chapter 8.

4.3 Classifier training

During training, the optimal classification model was constructed by iterating through the parameters of the classifier. Once this model is found, the regression function was computed for each of the remaining nodule candidates. The threshold applied to this value will split the candidates into positive and negative classes. Therefore, each operating point is described by: (a) optimal classification model, (b) threshold value on the corresponding regression function. In other words, if the detection system is to be evaluated for a certain operating point, one will need to apply both the model and the regression threshold to the evaluation set.

For the purpose of reporting the continuous set of performances (FROC curve) on the evaluation set, the regression threshold was varied and all obtained pairs of sensitivity and false positive rate per case were recorded.

In the training of the classifier, it is necessary to determine which parameter/model is the best, or results in the best detection performance. A classification model was considered the best if its sensitivity was highest across the entire range of false positive rates. Sometimes, two FROC curves corresponding to different classification models intersect; therefore, there was no way to

say which model is better. In this case, the model resulted in higher sensitivity at the rate of 1 FP/case was selected.

4.4 Measuring overall detection performance

Since both solid and nonsolid subsystems were designed and optimized for the detection of the nodules of their respective types, it may be necessary to evaluate overall performance by combining the output of the two. This can be done by taking the operating points from both "solid" and "nonsolid" FROC curves at the same sensitivity level S and finding corresponding classifier thresholds on the regression output τ_s and τ_n. If we denote the final set of solid nodule candidates as $\{c_i\}_{i=1,L}$, corresponding classifier regression function as $r_s(c_i)$, the final set of nonsolid nodule candidates as $\{\phi_j\}_{j=1,M}$, and the regression function as $r_n(\phi_j)$, the combined set of positive candidates would be:

$$\begin{aligned}\{C_k\}_{k=1,L+M} = &\{c_i : r_s(c_i) > \tau_s\}_{i=1,L} \\ &\cup \{\phi_j : r_n(\phi_j) > \tau_n\}_{j=1,M}.\end{aligned} \tag{4.2}$$

Resulting positive candidates are then assigned to either the $True$ or $False$ class using the procedure described above, regardless of nodule consistency.

By varying sensitivity level S from 0.0 to 1.0, one can obtain the set of operating points that would make up the resultant FROC curve.

Indeed, other strategies for obtaining combined FROC curves may be used. For example, one may weigh the importance of solid nodules relative to the importance of nonsolid nodules, or consider equal levels in the rate of false positives while establishing the correspondence across the operating points. These

strategies would result in alternative FROC curves; however, they are not explored within the scope of this thesis and are left for future consideration.

CHAPTER 5

MULTISCALE LAPLACIAN OF GAUSSIAN APPROACH TO CANDIDATE GENERATION

Candidate generation, the first stage for most computer aided detection (CAD) systems, rapidly scans the entire image data for any possible abnormality locations, while the subsequent stages of the CAD system refine the candidates list to determine the most probable or significant of these candidates. The candidate generator creates a list of the locations and provides a size estimate for each candidate. A multiscale scale-normalized Laplacian of Gaussian (LoG) filtering method for detecting pulmonary nodules in whole-lung CT scans, presented in this chapter, achieves a high sensitivity for both solid and nonsolid pulmonary nodules.

5.1 Introduction

The most commonly used architecture for a nodule CAD system is composed of the following stages: search space demarcation (usually segmentation of the lung field and/or identification of its boundary), nodule candidate generation within the search space, and reduction of false positive findings. This architecture is general enough to incorporate almost any work published in the field of pulmonary nodule detection. The only differences among published CAD systems are in the algorithms that are employed at each particular step and the final outcome of detection. For this reason, the different stages of a CAD system may be considered separately, but unfortunately, many authors do not report their performance as a separate item from total system performance.

In this chapter, the development and evaluation of only the nodule candidate generator in isolation from other stages is considered. The desired requirements for a nodule candidate generator are specified in the following.

Candidate generators determine the upper bound on CAD system sensitivity. Even though the resultant sensitivity of an automated detection system depends on each of its components, it cannot reach the value higher than the one of the generator. Therefore, the primary requirement for the candidate generator is that it accurately identifies the locations of as many nodules as possible, independent of their size and geometrical morphology, thus providing an initial sensitivity close to 1.0.

The second requirement is that the candidate generator provide an accurate size estimate for each candidate. This is critical for the subsequent stage of false positive elimination, where the feature computation algorithms and resultant feature values are often dependent on provided candidate size. Moreover, the evaluation of the performance of the detection system with respect to a minimum cut-off is more reliable when the nodule size estimation error is minimized.

Geometric positional accuracy is another important characteristic of a nodule candidate generator: the values of computed image features often depend on the geometric location of the candidate. If the centroid of a candidate is located far from the centroid of the nodule, the resultant features may have poor discrimination power. In addition to feature extraction, high positional accuracy is important for proper evaluation of the sensitivity. For example, if the system detects a vessel in close proximity to nodule, the decision should not be confused with the proper identification of the nodule and qualified as a false

positive.

Finally, the candidate generator must be computationally efficient. In commercial CAD settings, the constraint on the algorithm execution time is determined either by a customer, or, if systems with similar detection performance exist, by the market. For commercial lung CAD systems, the average processing time is on the order of 5-10 minutes [Das et al., 2006].

False positive rate, which is usually measured as the number of false findings produced by the detection algorithm per CT scan, is an important measure of overall CAD system performance. In the context of candidate generation, however, achieving a low false positive rate is secondary to the stated requirements and, therefore, is not considered in this chapter.

Example algorithms for nodule candidate generation and corresponding sensitivities evaluated on datasets containing at least 100 pulmonary nodules can be found in the works of Zhao [2003]; Farag et al. [2004]; Armato et al. [2005]; Enquobahrie et al. [2007]; Pu et al. [2008]; Li et al. [2008]; Murphy et al. [2009] and are summarized in Table 5.1. Reported sensitivities in these studies ranged from 88.3% to 98.7%. Unfortunately, the accuracy of size estimation and localization were not explored by the authors.

Zhao [2003] used the multiple-level thresholding that involves segmenting the lungs, setting an image intensity threshold at a certain value and detecting connected three-dimensional objects. The threshold is then decreased in a step-wise manner recovering less-dense objects and extending the objects detected at higher thresholds. The decision whether to treat a connected object as a nodule candidate is based on a set of rules that incorporate size, density, and

Table 5.1: Examples of reported candidate generators with their initial sensitivities.

First author	Algorithm	Nodules	Sensitivity	False positive rate
Zhao, 2003	Local density maximum	266, $\geq$ 2 mm	0.944	906
Farag, 2004	Template matching	130	0.846	49 (per slice)
Armato, 2005	Multiple gray-level thresholding	470	0.883	539
Enquobahrie, 2007	Binary image features	499, $\geq$ 4 mm	0.958	3785
Pu, 2008	Signed distance field	184, $\geq$ 4 mm	0.951	1200
Li, 2008	Selective enhancement filters	153, $\geq$ 4 mm	0.987	140
Murphy, 2009	Local image features	3981, $\geq$ 3 mm	0.976	700

a variable describing the acceptable change of an object's volume. There are additional parameters that specify intensity threshold bound and step size. Another multiple intensity level thresholding technique was used by Armato et al. [2005]. Here, instead of decreasing threshold, the authors increased it stepwise and recorded all the objects smaller than a certain size. The algorithm described by Farag et al. [2004] involves template matching of the local lung regions with predefined nodule templates of various size and shape using normalized cross-correlation measure. Since the dimensionality of the search space was high, the authors did not perform an exhaustive search, but instead, resorted to a genetic algorithm, that provided the position and initial size of the candidate templates. Pu et al. [2008] first obtained a binary image of a CT scan using an empirically selected threshold. Nodule candidate generation was quite simple — the authors calculated signed Euclidean distance fields and recorded local maxima and corresponding distances as nodule candidates. The work of Li et al. [2008] relies on the use of filters for nodule enhancement and suppression of normal anatomic structures such as blood vessels and lung walls. These filters are multiscale in nature and are able to characterize the local image structures such as sphere, cylinder or plane. This discrimination is possible due to computing local image curvature using second-order image derivatives. After the original CT scan is filtered, thresholding identifies nodule candidates. In the work of Murphy et al. [2009], nodule candidates were identified by computing shape index and curvedness for each image location. Voxels having these values above a certain threshold were clustered together and adjusted to form nodule candidates. The method described in Enquobahrie et al. [2007] is capable of identifying solid nodules and consists of two modules: for detecting isolated (inside lung parenchyma attached to small pulmonary vessels and airways) and at-

tached (adjacent to lung wall, mediastinum surface) nodules. Briefly, the first module works with the thresholded image of the lungs and looks for pixel clusters that have a certain minimum size and limited extent. The second module is based on morphological processing and analysis of the lung boundary shape for identifying protrusions of significant size. Both modules have an associated set of parameters that need to be tuned to achieve maximum performance. Since this method was previously developed and used in our research group, it will be used as a reference for the experiments described further in this chapter.

All of the candidate generation techniques mentioned here have either thresholds or parameters that need to be fine-tuned to achieve the best performance. Even though such flexibility may result in very high sensitivity, optimization of these generators to datasets with different image acquisition parameters or target nodule size ranges may require additional work. In addition, the majority of previously developed candidate generators rely on correct segmentation of the lung volume that is critical for proper localization of nodules adjacent to lung wall. In order to minimize the effect of these issues, in the design of the nodule candidate generator the number of control parameters was minimized to make the generator less dependent on the outcome of the lung segmentation.

The candidate generator presented in this paper is based on multiscale Laplacian of Gaussian (LoG) filtering. LoG filtering has been used for multiscale analysis in many computer vision and image analysis applications. Use of LoG filtering for detecting edges on digital images was first proposed by Marr and Hildreth [1980]. Subsequently, LoG filtering was used for enhancing image ”blobs” for locating the aorta in MR imagery by Jiang and Merickel [1989];

and for identifying image texture elements by Blostein and Ahuja [1989]. The method of detecting ”blobs” as scale-space structures was formalized by Lindeberg [1993] and later employed by many authors. LoG-based ”blob” detection has been used for nodule detection in two-dimensional chest radiographs by Schilham et al. [2003]. The application of LoG filtering for localization of nodules in the three-dimensional CT scans was first proposed by Reeves et al. [2006] and evaluated with respect to the nodule size estimation properties in the works of Jirapatnakul et al. [2009] and Diciotti et al. [2010]. The previous application of LoG filtering with respect to computerized nodule detection from CT has been limited to the work of Dolejsi and Kybic [2007].

5.2 Method

The presented method takes into account the distinction between different nodule types. In the following section two fundamentally different classes of nodules are described and a specific approach tailored to each nodule type is then developed.

5.2.1 Nodule types dichotomy

A solid pulmonary nodule is an approximately spherical lesion having an image intensity similar to that of soft tissue (median = -294 HU, standard deviation = 164 HU as found by Browder [2007]). This compares to the intensity of air which, by definition, has a value of -1000 HU. The size range for pulmonary nodules is from 3 mm to 3 cm in diameter [Austin et al., 1996]. The

majority of the image intensity in the lungs is the aerated lung parenchyma which has an image intensity approaching that of air (median = -810 HU, stdev = 63 HU [Browder, 2007]). Other normal visible structures in the lungs are airways and blood vessels. These are branching cylindrical structures that have similar image intensity to soft tissue. The airways also have a lumen that is the intensity of air. The minimum solid nodule diameter to consider radiological lung examination non-negative, varies in the range of 4.5 - 5 mm [Henschke et al., 1999; Xu et al., 2006; Lopes Pegna et al., 2009].

The nonsolid nodule is a second class of pulmonary nodule that, unlike the solid nodule, is caused by layers of cells lining the alveoli and the airways. In overall appearance it has a density that is slightly more than that of the lung parenchyma since there is still a significant amount of air within the airways and alveoli (median = -680 HU, standard deviation = 58 HU [Browder, 2007]). In these lesions, more dense vessels and larger airways may be visible. In general, these lesions are also approximately spherical. Nonsolid nodules are harder to see compared to solid nodules and when small in size are difficult to discriminate from other parenchyma density variations. For this reason clinical significance is usually considered for nonsolid nodules having a minimum size in the range 8 - 10 mm.

Typical pulmonary nodules of both types are shown in Figure 1.2. Nodules may have variation in size, shape, intensity profile and attachment morphology.

In the following section the model of a solid nodule and the corresponding detection method are presented. The solid nodule model is then extended to deal with the more complex case of nonsolid nodules.

5.2.2 Solid pulmonary nodule image model

The objective of the candidate generator is to identify pulmonary nodules from other normal structures in the lung region that have a density greater than the lung parenchyma background. For this reason, a solid pulmonary nodule is modelled as a solid spherical structure with the intensity of soft tissue (equal to 1) on a background having the same intensity as lung parenchyma (equal to 0). First, the properties of such a model are explored in isolation from other objects. Then, the applicability of the model is investigated in the presence of interference from the structures resembling pulmonary vessels and chest walls.

The Laplacian of Gaussian has been described as a detector that responds to ”bright regions on dark background” [Lindeberg, 1993]. That is the general definition that is used for the term ”blob” in this paper. Marr in his original work [Marr, 1976; Marr and Hildreth, 1980] more formally defined a ”blob” as a primitive compact image element enclosed in the closed contour made of the LoG zero-crossings.

5.2.3 Blob detection as scale-space normalized LoG filtering

A method that detects the location and estimates the scale of blob-like structures is described by Lindeberg [1998]. Identification of ”blobs” is accomplished by finding scale-space maxima and minima of scale-normalized Laplacian

$$\nabla^2_{norm} L(X, \sigma) : \Re^3 \times \Re \rightarrow \Re. \tag{5.1}$$

Here $L(X, \sigma)$ is the scale-space representation obtained for image $I(X)$ by convolving it with Gaussian kernel $G(X, \sigma)$ at continuous set of scales:

$$L(X, \sigma) = G(X, \sigma) * I(X). \tag{5.2}$$

In this chapter, the following notation is used for a three-dimensional coordinate vector: $X = (x, y, z)$; and the variance of Gaussian kernel σ is referred to as the "scale parameter" or, simply, "scale" or "size."

The response function $\nabla^2_{norm} L(X, \sigma)$ may be expressed as follows using the properties of convolution:

$$\nabla^2_{norm} L(X, \sigma) = \nabla^2_{norm} G(X, \sigma) * I(X). \tag{5.3}$$

This operation is effectively the filtering of the original image with scale-normalized LoG kernels of continuously changing scale parameter.

Normalization is necessary to eliminate the effect of decreasing spatial derivatives with the increase of scale and is defined as the negated multiplication of LoG by σ^2:

$$\nabla^2_{norm} G(X, \sigma) = -\sigma^2 \nabla^2 G(X, \sigma). \tag{5.4}$$

By introducing the negative sign for normalization, the computations are brought to the domain of positive real numbers, where "bright" blobs are identified as local maxima of $\nabla^2_{norm} L(X, \sigma)$ instead of local minima. Later in the chapter, the negated scale-normalized LoG will be referred to as simply the "normalized LoG."

The initial set C of nodule candidates is constructed as a subset of local maxima of the response function $\nabla^2_{norm} L(X, \sigma)$, such that their spatial component is

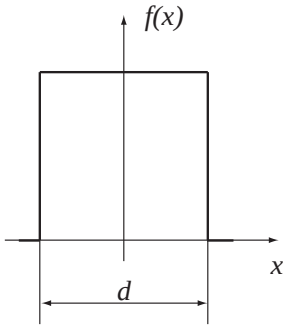


Figure 5.1: Rectangular function as one-dimensional representation of the nodule model.

located within the lungs:

$$C = \{(Y, \sigma) : Y \in Lungs, \\ (Y, \sigma) \in \{localmax\ \nabla^2_{norm} L(X, \sigma)\}\}. \tag{5.5}$$

5.2.4 Properties of the normalized LoG filter

To illustrate the concept of normalized LoG filtering, the nodule model is reduced to a one-dimensional representation — rectangular function as sketched in Figure 5.1.

In Figure 5.2 the responses of the normalized LoG filter to the rectangular function of fixed width are observed. An illustration of the filtering applied to an image having three rectangles of varying widths and intensities is shown in Figure 5.3. The maximum response occurs when the LoG kernel is located at the center of the rectangle and when the width of rectangle matches the size of its central lobe: $\sigma = d/2$. Linearity of the LoG filter causes the response to be proportional to the height of the rectangle.

In the three-dimensional domain the principle of multiscale filtering remains

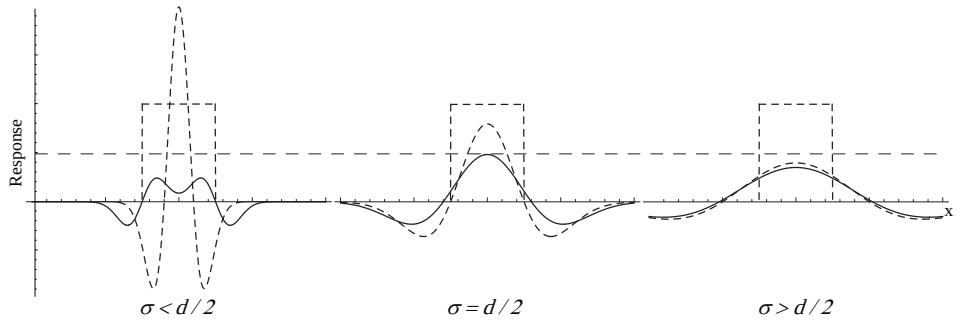


Figure 5.2: Illustration of multiscale LoG filtering: normalized LoG kernels of varying scales (curvy dashed line) are convolved with the rectangle function (dashed line). Maximum response (solid line) is observed when the size of the kernel corresponds to the width of the rectangle (i.e. normalized LoG zero-crossings coincide with the ″boundary″ of the rectangle). Rectangular function as one-dimensional representation of the nodule model.

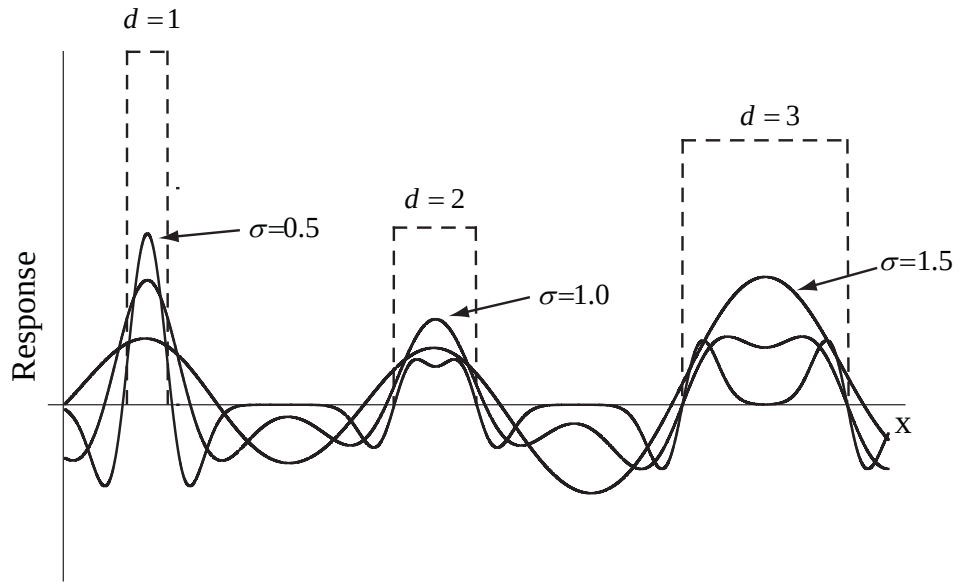


Figure 5.3: Illustration of multiscale LoG filtering with respect to three rectangular functions of different size and intensity. Three curves represent responses of the differently sized normalized LoG kernels. Maximum response is proportional to the height of the rectangle and is achieved when the kernel width corresponds to its size.

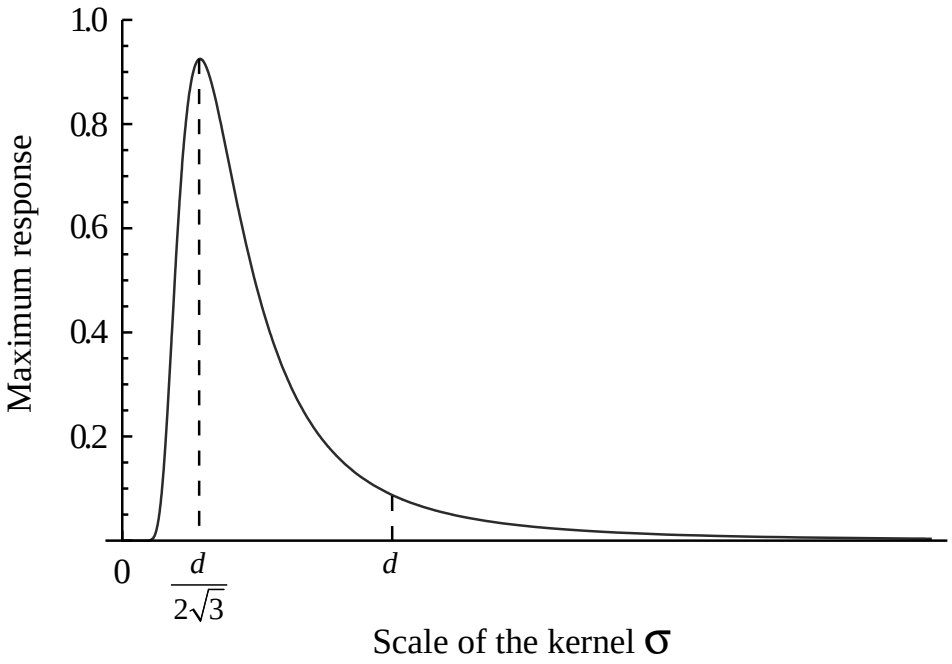


Figure 5.4: Response of the normalized LoG kernels with different scales σ to solid sphere of diameter d: local maximum is achieved at the value of $\sigma = d/2\sqrt{3}$.

the same. The responses of the normalized LoG filter with different scales σ to a solid sphere are shown in Figure 5.4. Maximum response is reached when the scale of the kernel σ corresponds to the diameter of the sphere d: $\sigma = d/2\sqrt{3}$.

5.2.5 Multiscale normalized LoG filtering

Theoretically, $\nabla^2_{norm} L(X, \sigma) : \Re^3 \times \Re \rightarrow \Re$ is a real function of four continuous variables: three spatial and one scale; however, for the implementation it is also necessary to specify an appropriate parameter quantization scheme. In our research, the quantization of spatial dimensions was selected to be identical to the quantization of the original image, i.e. the accuracy of the candidate centroid localization in each direction was limited to the voxel spacing interval along

corresponding axes.

As for the scale quantization, it is effectively the computation of $\nabla^2_{norm} G(X, \sigma) * I(X)$ for different values of σ, or "multiscale filtering." The normalized LoG response may thus be expressed as the function of discrete variables:

$$\nabla^2_{norm} L(X, \sigma) : S^3_X \times S_\sigma \rightarrow \Re, \tag{5.6}$$

where $S^3_X = \{0, 1, ..., x_{\max}\} \times \{0, 1, ..., y_{\max}\} \times \{0, 1, ..., z_{\max}\}$ is the discrete set of three-dimensional image coordinates and $S_\sigma = \{\sigma_i\}$ is the discrete set of scales. By adding the $Lungs$ spatial constraint, the search space for nodule candidates is obtained as $(S^3_X \cap Lungs) \times S_\sigma$.

As multiple convolutions need to be computed, the scale quantization strategy is of high importance and will affect both multiscale filter response and estimated nodule sizes. In addition, the processing of too many scales may result in unnecessary computational burden without any benefits for detection.

In the following section we determine the design bounds for a multiscale normalized LoG filter for detecting solid spheres in a noise free situation in which both soft tissue and lung parenchyma have constant intensity values. This will set a lower bound on the parameters that will be needed for a real application in which there is image noise and some variation in tissue intensities.

5.2.6 Scale quantization

A bound on maximum reduction in filter response

The maximum response of a scale-normalized LoG filter of scale σ to the solid sphere S_d of diameter d and unit intensity is reached at its center and can be found as

$$\begin{aligned} R(\sigma, d) &= \int\limits_{X \in S_d} \nabla^2_{norm} G(X, \sigma)\, dX = \\ &= \frac{d^3}{2^{2.5}\pi^{0.5}\sigma^3} \exp\left(\frac{-d^2}{8\sigma^2}\right). \end{aligned} \tag{5.7}$$

It can be shown, that for a given diameter d of the sphere, the magnitude of the response function is maximal and independent of sphere diameter:

$$R_{peak} = \sqrt{\frac{54}{\pi}} \exp(-1.5) \approx 0.925. \tag{5.8}$$

Let us consider two adjacent scales σ_1 and σ_2 ($\sigma_1 > \sigma_1$) of the multiscale filter. The sphere diameters corresponding to maximum response will be $d_1 = 2\sqrt{3}\sigma_1$ and $d_2 = 2\sqrt{3}\sigma_2$, respectively. These spheres will result in peak responses of the multiscale filter:

$$R(\sigma_1, d_1) = R(\sigma_2, d_2) = R_{peak}, \tag{5.9}$$

as shown in Figure 5.5.

As the scale space is quantized, the value of maximum response will never reach R_{peak}, unless the diameter perfectly corresponds to a scale in the quantized set. It will result in spheres of different sizes having filter responses that diverge from the peak value. Any sphere of diameter d, such that $d_1 < d < d_2$, would result in a smaller filter response. Let us estimate the diameter d_{dip} that results in a minimal response R_{dip} of the filter to a spherical model. Clearly, the minimum

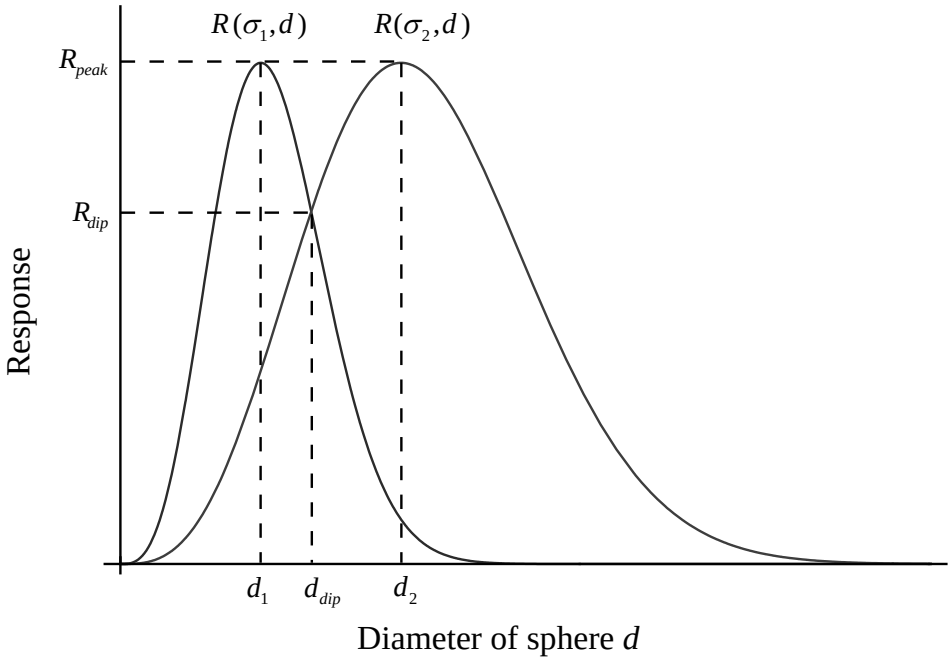


Figure 5.5: Response of the normalized LoG kernels of scales σ_1 and σ_2 to the solid sphere of unit intensity. Spheres with diameters $d_1 = 2\sqrt{3}\sigma_1$ and $d_2 = 2\sqrt{3}\sigma_2$ result in peak responses; differently sized spheres result in lower response.

value will be reached when the responses from both scales are the same:

$$R_{dip} = L(\sigma_1, d_{dip}) = L(\sigma_2, d_{dip}). \tag{5.10}$$

The nontrivial solution to this equation with respect to d_{dip} is:

$$d_{dip} = \sqrt{8}\sigma_1\sigma_2\sqrt{\frac{\ln {\sigma_2}^3 - \ln {\sigma_1}^3}{{\sigma_2}^2 - {\sigma_1}^2}}, \tag{5.11}$$

which results in:

$$R_{dip} = \frac{4}{\sqrt{\pi}}\left(\sigma_1/\sigma_2\right)^{\frac{3}{1-(\sigma_1/\sigma_2)^2}}\left(\frac{3\ln\left(\sigma_1/\sigma_2\right)}{\left(\sigma_1/\sigma_2\right)^2 - 1}\right)^{1.5}. \tag{5.12}$$

The minimal response depends on only the ratio of the adjacent scales. This means that quantization with the scale increasing in a geometric progression will result in fixed bounded error of the filter response. Therefore, to maintain a constant error bound over the range of sphere sizes of interest, we establish a

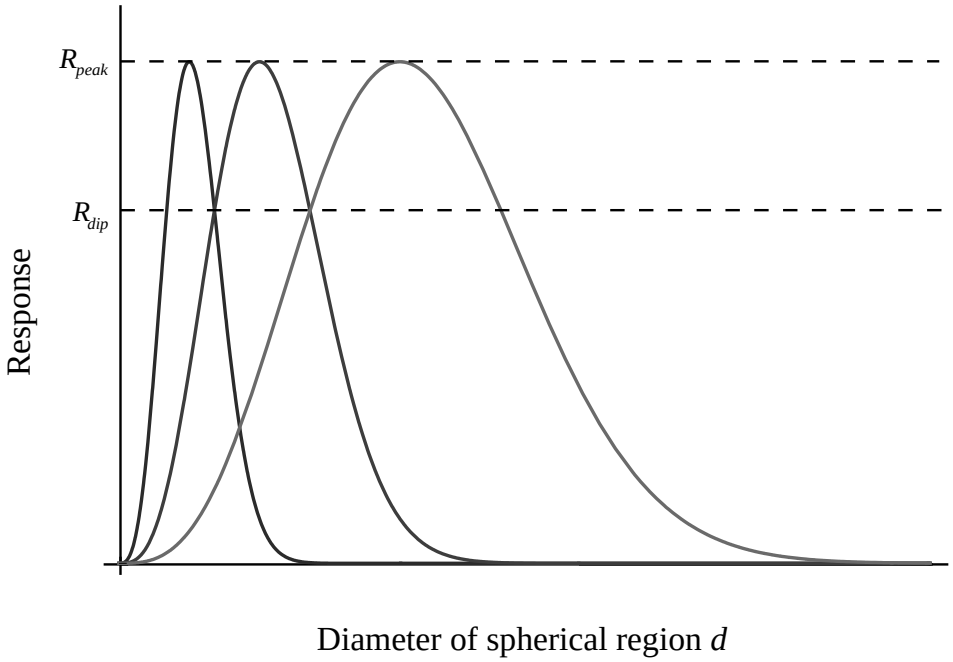


Figure 5.6: Exponentially increasing scale $\sigma_{i+1} = k\sigma_i$ results in the reduction in response bounded from below by R_{dip}.

set of scales with a geometric progression with a step size of k with respect to the smaller scale:

$$\sigma_{i+1} = k\sigma_i. \tag{5.13}$$

For such a scale set the scale is increased exponentially, which results in an R_{dip} that is independent from the scale as illustrated in Figure 5.6.

A bound on filter response for shape confusion

One of the criteria for selecting an appropriate scale quantization is the ability of the filter to discriminate between basic geometrical shapes. Given that pulmonary vessels are the other common structure within the lungs, we consider the behavior of a normalized LoG filter on a such structure, which is modelled as a sufficiently long solid cylinder.

The maximal response of the normalized LoG filter to a solid cylinder C_d of diameter d is given by

$$\begin{aligned} R'(\sigma, d) &= \int_{X \in C_d} \nabla^2_{norm} G(X, \sigma)\, dX \\ &= \frac{d^2}{4s^2} \exp\left(\frac{-d^2}{8\sigma^2}\right). \end{aligned} \tag{5.14}$$

This response reaches its maximal value of $R'_{peak} = 2/e \approx 0.736$, when $\sigma = d/2\sqrt{2}$.

The graph in Figure 5.7 shows how the selection of the coefficient k affects the reduction in response due to scale quantization: as the value of k approaches 1.0, the value of R_{dip} approaches $R_{peak} \approx 0.925$ (shown as top dashed line). This is the maximum response that can be achieved by convolving a normalized LoG kernel with a solid sphere. The bottom dashed line corresponds to the maximum response of $R'_{peak} \approx 0.736$ obtained for a solid cylinder. As the graph shows, if the scale quantization is too rough ($k > 1.746$), the multiscale filter will not be capable of reliable discrimination between solid sphere and a solid cylinder.

A bound on size estimation error

Scale quantization has an important impact on candidate size measurement accuracy. To estimate this impact, let us consider the worst case scenario, when a solid sphere has a diameter that approaches the value of d_{dip} corresponding to the maximum reduction in filter response.

The largest diameter underestimation (UE) will be reached when the diameter reaches d_{dip} from the left side as shown in Figure 5.8. The relative size

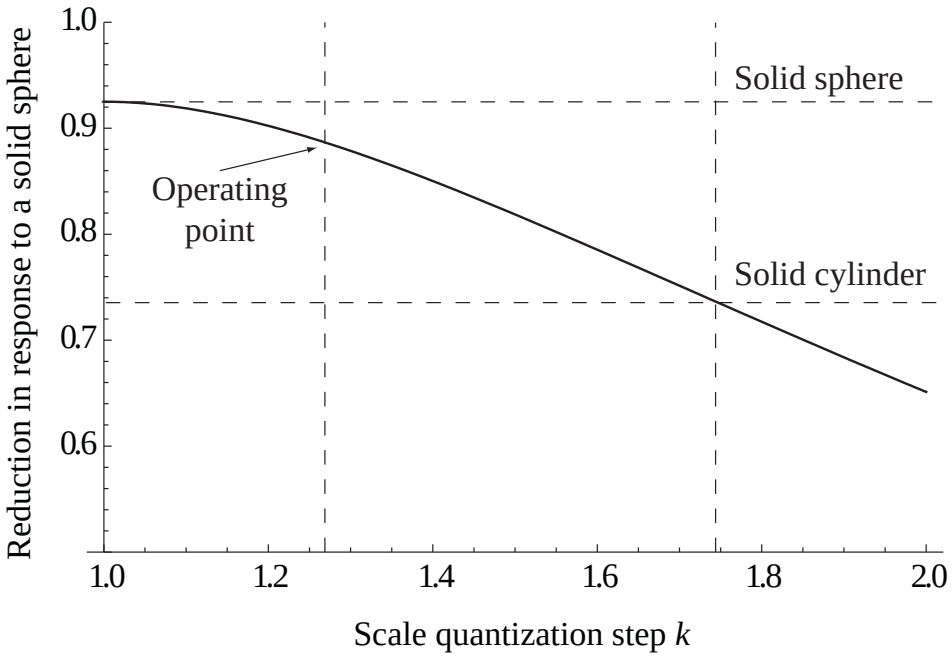


Figure 5.7: Selection of operating quantization step. Solid curve shows the reduction of filter response R_{dip} of the normalized LoG filter response to the solid sphere with increased quantization step k. Dashed lines shows the maximum responses to a solid sphere and a cylinder.

measurement error will be equal to:

$$\delta d_{ue} = \frac{d_{dip} - d_1}{d_{dip}} = 1 - \sqrt{\frac{1 - k^{-2}}{2 \ln k}}. \tag{5.15}$$

The largest overestimation (OE) error, is when the diameter approaches d_{dip} from the right side:

$$\delta d_{oe} = \frac{d_2 - d_{dip}}{d_{dip}} = \sqrt{\frac{k^2 - 1}{2 \ln k}} - 1. \tag{5.16}$$

Due to the asymmetry of response function (5.7), the overestimation error is always larger than the underestimation error as shown in Figure 5.9. On the one hand, the values of k close to 1 result in both less filter confusion and better size estimation. On the other hand, with such a small step size, the number of scales needed to cover the nodule size range of interest will be high, which is not desirable from the computational cost point of view.

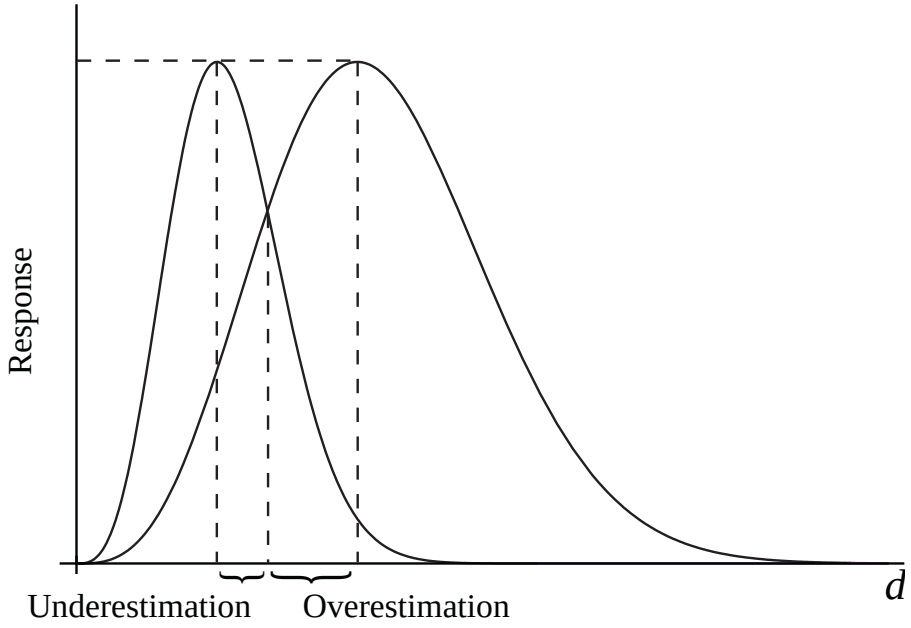


Figure 5.8: Underestimation and overestimation of the sphere size due to scale quantization.

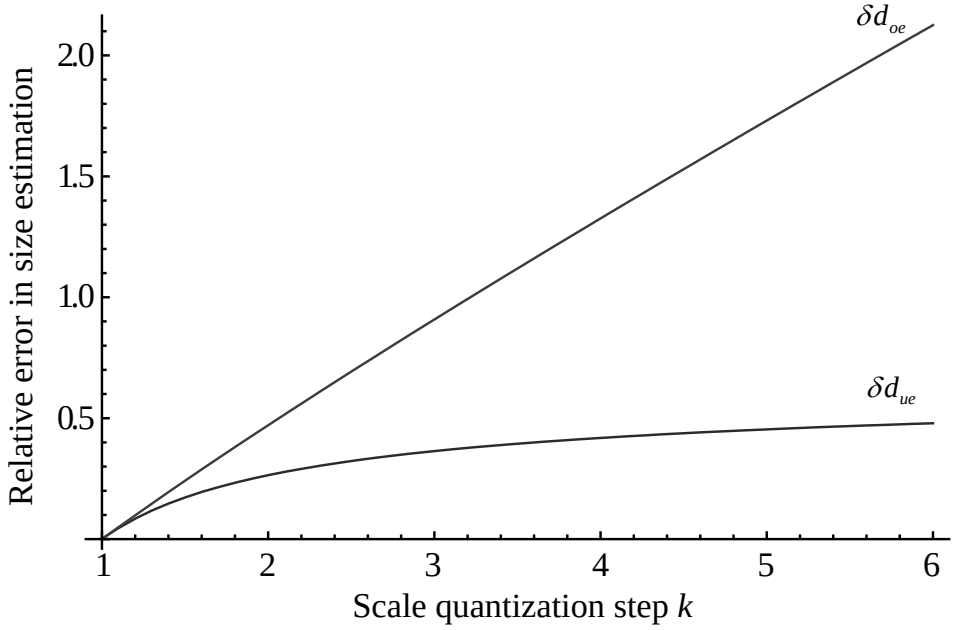


Figure 5.9: Relative error in solid sphere size overestimation (OE) and underestimation (UE) with respect to size quantization step k.

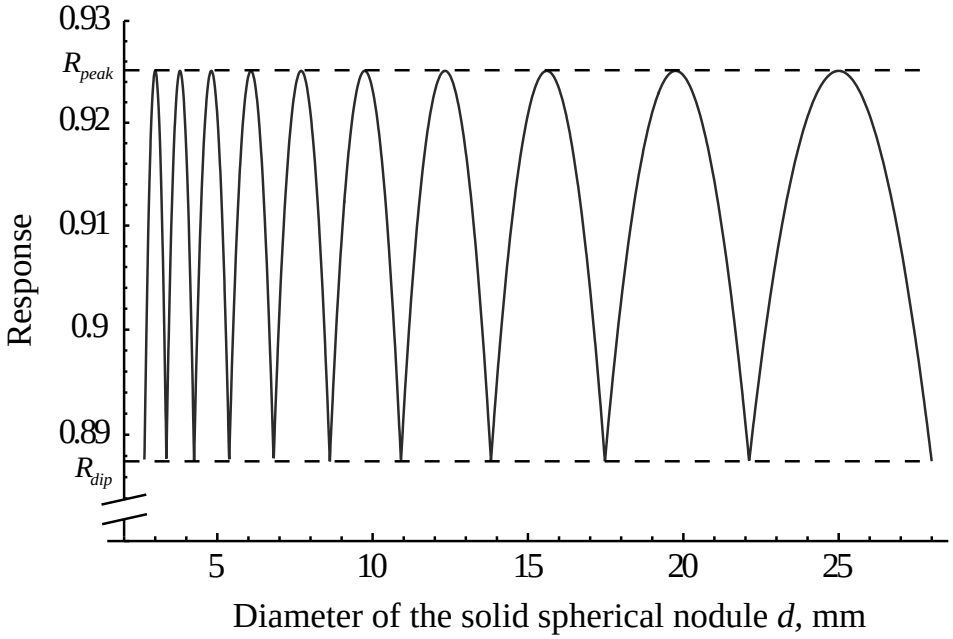


Figure 5.10: Responses of the multiscale LoG filter to three-dimensional solid spheres of different diameters.

In our experiments, the value of $k = 1.27$ and 10 scales corresponding to sphere diameters increasing in geometrical progression from 3 to 25 mm were used. This was done to match the size range of targeted nodules plus two boundary scales. The main reasons for selection of this quantization strategy were maintaining a bound on reduction in LoG filter response and a bound on shape confusion. In this configuration, the highest reduction in response results in $R_{dip} \approx 0.887$. Relative errors in size underestimation and size overestimation were no greater than $\delta d_{ue} = 0.11$ and $\delta d_{oe} = 0.13$, respectively. The set of quantized diameters d_i, corresponding scales σ_i, and effective diameter ranges are shown in Table 5.2. All spheres in each $i - th$ effective diameter range (fourth column) will be detected as having diameter d_i.

Corresponding response function of the multiscale LoG filter is illustrated in Figure 5.10.

Table 5.2: Quantization of candidates sizes. The columns are: scale index, diameter of the kernel, corresponding scale and range of the candidate sizes assigned to the scale.

i	d_i, mm	σ_i	$range_i$, mm
0 (boundary scale)	2.37	0.68	N/A
1	3.00	0.86	2.65 - 3.35
2	3.79	1.09	3.35 - 4.25
3	4.80	1.38	4.25 - 5.38
4	6.08	1.75	5.38 - 6.81
5	7.69	2.22	6.81 - 8.62
6	9.74	2.81	8.62 - 10.91
7	12.33	3.55	10.91 - 13.80
8	15.60	4.50	13.80 - 17.47
9	19.75	5.70	17.47 - 22.11
10	25.00	7.21	22.11 - 27.99
11 (boundary scale)	31.64	9.13	N/A

5.2.7 The impact of spatial interference

In previous sections, the scale quantization effects and properties of our nodule model in isolation from other pulmonary structures were reviewed. For a better understanding of the behavior of the LoG filtering in practical sense, it is important to estimate the stablity of the nodule model in the presence of other spatial structures such as solid cylinder and wall that closely approximate a pulmonary

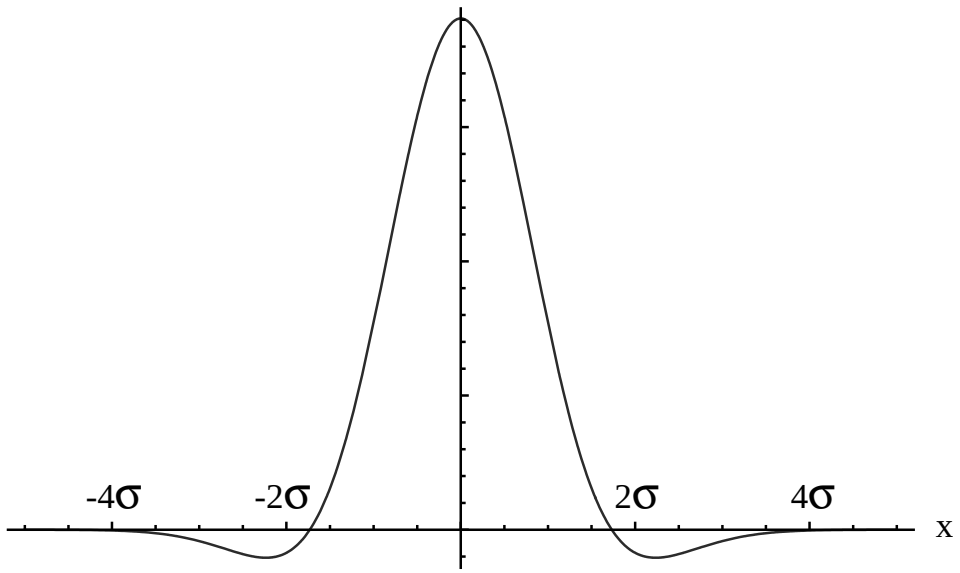


Figure 5.11: Central slice of the three-dimensional normalized LoG kernel centered at the origin of coordinate system. Major part of the volume of support is enclosed within the radius of 4σ from the origin.

vessel and chest wall, respectively.

For example, if a nodule is located close to a pulmonary vessel or "attached" to a chest wall, the response of the filter to the nodule would be altered due to superposition in the response to the interfering structure. This effect is amplified as the objects get closer to each other. From the profile of the negated LoG kernel centered at the origin of the coordinate system, shown in Figure 5.11, one may infer that all significant nonzero values are concentrated near the origin and do not extend beyond 4σ (or slightly greater than doubled zero-crossing) distance. This means that the objects located closer than 4σ from each other may produce a noticeable interference effect.

To quantify the interference effect, two simulations for an observed solid sphere model of unit diameter and intensity were conducted. The purpose

of these simulations was to imitate the most common interference happening within the lungs: with a pulmonary vessel and with a chest wall. In the first simulation we recorded how a proximal long solid cylinder of the same diameter and intensity affects the sphere filter response and size estimate. In the second simulation the solid sphere was placed near a solid flat wall of sufficiently large size and thickness. The simulations were conducted using discrete synthetic three-dimensional images of these objects. For each distance (measured in sphere diameters), the normalized LoG space maxima corresponding to the sphere was found and the best scale (size estimate) and filter response were recorded.

Results of the sphere-cylinder interference simulation are shown in Figure 5.12. When the distance from the central axis of the cylinder to the center of the sphere was greater than 1.5, no negative effects from interference were observed. As the distance between the objects is reduced, the response of the filter decreases. The size estimate maintained close to 1.0 until the objects started to form an overlap. Additional decreasing of the distance to 0.5 resulted in a decrease of response. Finally, after passing the 0.5 threshold, the normalized LoG filter was not able to distinguish between the sphere and the cylinder and detected a single ”blob” instead; a jump in both response and size estimate was observed. Finally, when the objects overlay entirely, the filter ”converges” to the response and size estimates of an ideal standalone cylinder. The effect of sphere-wall interference is shown in Figure 5.13. Here, as the sphere approaches and merges with the wall, its estimated size is decreased. Even though there is no unambiguous definition of the ”size” for a sphere partially attached to a wall, it is reasonable to assume that it should be smaller than the one of the isolated sphere. Similarly, as the sphere is ”immersed” into the wall, the response de-

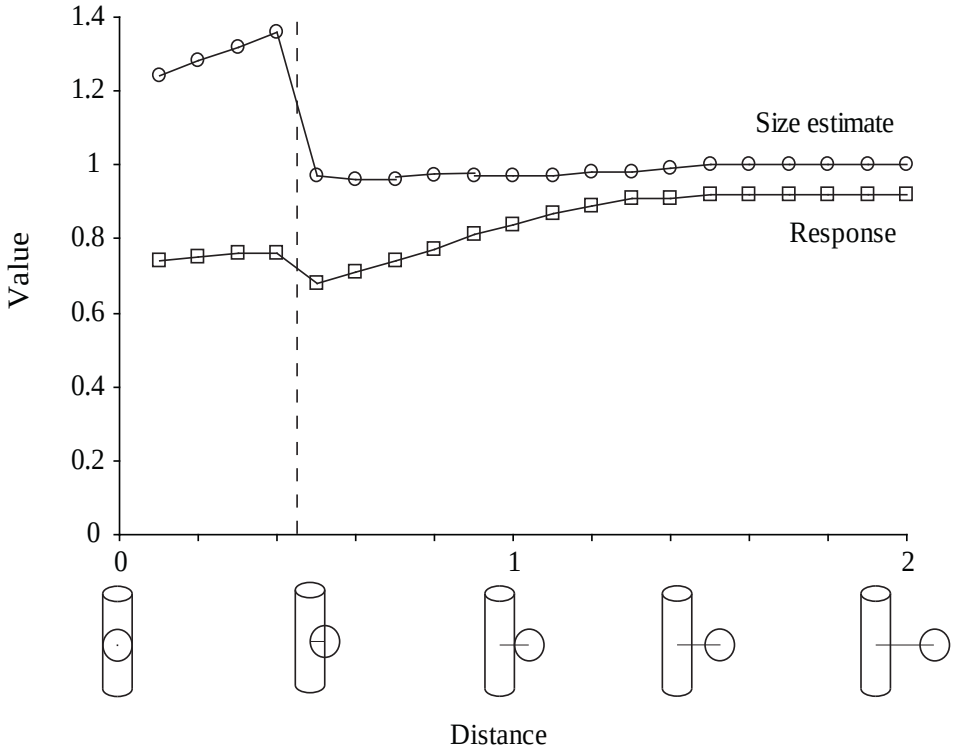


Figure 5.12: Cylinder-sphere interference. Effect of the distance between the objects on filter response and sphere size estimate. Dashed line indicates the border between detecting one single ”blob” and two separate ”blobs.”

creases in agreement with an increased degree of attachment.

In both scenarios, the sphere, unless merged with the interfering object entirely, resulted in a distinct maximum in the normalized LoG search space. In spite of diminished response and altered size for proximity distances less than 0.5, this can be resolved by a later stage of a CAD system that may still classify this candidate as a true positive.

These simulations confirmed that the interference exists and affects the filter response and size estimate of proximal objects. The simulations also showed that the interference effect is within reasonable limits and should not cause ma-

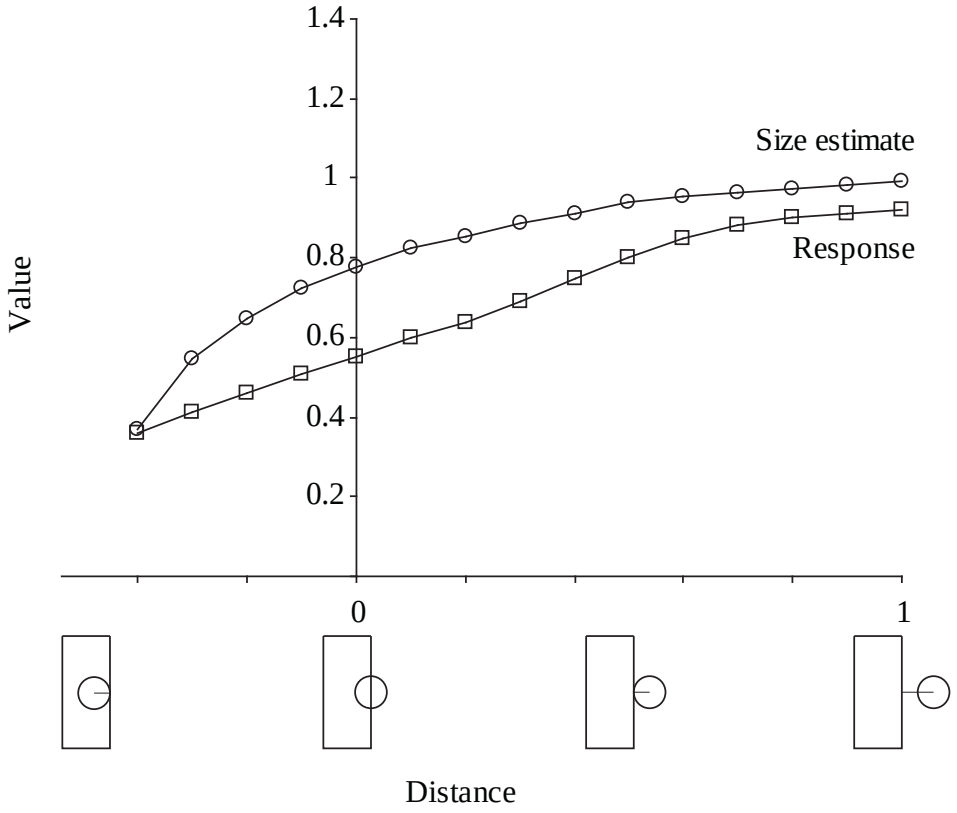


Figure 5.13: Sphere-wall interference. Effect of the distance between the centroid of a unit sphere and the wall on the filter response and size estimate.

jor obstacles to candidate identification. However, if the adjacent object has an intensity that is higher than the target nodule, the interference effect may be amplified, e.g. in the case of detecting nonsolid nodules in proximity to solid structures. This is discussed later in the nonsolid candidate generation subsection.

5.2.8 Minimum response criterion

The proposed multiscale LoG filtering scheme would produce local maxima in the search space for both true nodules and other image structures including

noise. As a post-processing step we propose to threshold and eliminate the low-response candidates that do not meet our solid nodule model, both shape- and intensity-wise.

To determine mimimum response threshold it is necessary to consider the "worst" scenario where a true nodule would have a minimum possible response. This may happen when the response to the nodule is degraded due to all of the following factors all together: (a) interference from other pulmonary structure; (b) quantization of detection scales; (c) low tissue contrast.

To model a situation such as this, let us refer to the cylinder-sphere interference simulation experiment, described above. The greatest reduction in response due to interference results in the response for sphere equal to: $R_{int} \approx 0.7$ as shown in Figure 5.12. If we consider the reduction in response due to quantization, the "worst" response will be further reduced by a factor of: $R_{dip}/R_{peak} = 0.958$. Here we made an assumption that the reduction of response due to quantization in the case of interference is of the same order as the one computed for an isolated sphere.

Previously we considered the objects having intensity of 1.0 and backround having intensity of 0.0. Using the linearity of the LoG filtering, one can recompute the response function in the real CT scan domain. The median intensity for lung parenchyma tissue is I_{paren} = -810 HU [Browder, 2007] and the lower bound on solid tissue intensity is $I_{solid,\min}$ = -474 HU [Browder et al., 2007]. Therefore, the final "worst" response can be written as

$$R_{CT} \approx R_{int} \frac{R_{dip}}{R_{peak}} \cdot (I_{solid,\min} - I_{paren}) = 226. \tag{5.17}$$

In other words, the scenario where solid nodule of lowest contrast is affected by both maximal spatial interference and the extreme scale quantization would

result in a LoG filter response of 226. The values of responses greater than this threshold value are sufficient for detecting solid nodules and filtering out noise and structures other than nodules.

One may hypothetically think of an even worse situation where a sphere is merged with a solid wall of greater intensity. This situation may happen if there is a rib that is close to a large nodule "attached" to the chest wall. In this case, the response threshold must be lowered, or, alternatively, the high intensity rib may be suppressed from the image; this technique is considered later in the chapter.

5.2.9 Candidate generation scheme

The main steps for the multiscale normalized LoG solid nodule candidate generator are the following:

1. Segment the spatial mask $Lungs$ from the CT image.
2. Compute $\nabla^2_{norm} L(X, \sigma_i)$ response function for discrete set of scales.
3. Identify nodule candidates from given $Lungs$ and $\nabla^2_{norm} L(X, \sigma)$.
4. Delete candidates with low filter response.

The first step of the algorithm involves computation of the lungs spatial mask that will limit the spatial search space of the algorithm. It was obtained as described in the search space demarcation section of Enquobahrie et al. [2007]. Prior to candidate generation, the lungs region was extended outwards by morphological dilation with a solid sphere of 10 mm. The purpose of such an extension was to account for lung segmentation imperfections and to make sure that the segmented lungs will encompass all nodules.

The main purpose of the second step is to compute the response function $\nabla^2_{norm} L(X, \sigma_i)$ for a discrete set of scales. To optimize the convolution operation, the convolution theorem is used, while the computation is carried in the frequency domain. An outline of the optimization is given in subsection 5.2.10 while the scheme of the entire step is shown in Algorithm 1. In short, to find the response for each of the scales, the Discrete Fourier Transforms (DFT) of the original image was found and multiplied with a pre-computed transform of the normalized LoG function; then the inverse DFT was taken and normalized. For computation of the DFTs, the FFTW library [Frigo and Johnson, 2005] was used. It allows computation of the convolution of two 512x512x512 images within a few seconds.

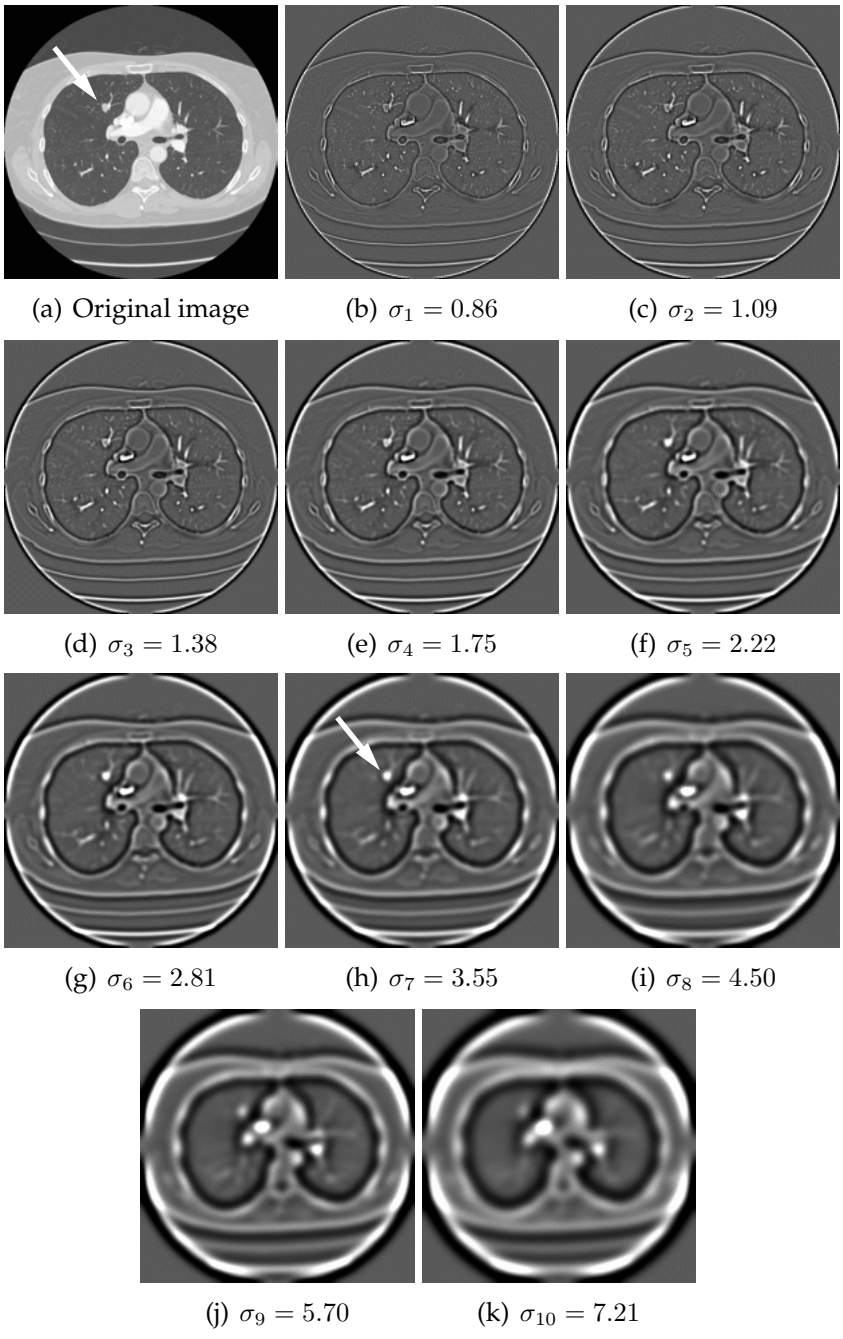
Algorithm 1: Constructing discrete response function

S_σ	discrete set of scales
$I(X)$	original image

$I_{\mathcal{F}}(\Omega) \leftarrow \mathcal{F}\{I(X)\}$	find the DFT of original image
for each $\sigma_i \in S_\sigma$	for each discretization step
$M(\Omega, \sigma_i) \leftarrow I_{\mathcal{F}}(\Omega) \cdot \mathcal{F}\left\{\nabla^2 G(X, \sigma_i)\right\}$	multiply the image and LoG DFTs
$\nabla^2 L(X, \sigma_i) \leftarrow \mathcal{F}^{-1}\left\{M(\Omega, \sigma_i)\right\}$	find the inverse DFT
$\nabla^2_{norm} L(X, \sigma_i) \leftarrow -\sigma_i^2 \nabla^2 L(X, \sigma_i)$	normalize
end for	

The response function for the example CT image is shown in Figure 5.14. It consists of multiple response functions computed for each scale.

The third step involves finding the candidates given the set of search subspaces for each scale. Algorithm 2 illustrates the procedure.



(a) Original image (b) $\sigma_1 = 0.86$ (c) $\sigma_2 = 1.09$

(d) $\sigma_3 = 1.38$ (e) $\sigma_4 = 1.75$ (f) $\sigma_5 = 2.22$

(g) $\sigma_6 = 2.81$ (h) $\sigma_7 = 3.55$ (i) $\sigma_8 = 4.50$

(j) $\sigma_9 = 5.70$ (k) $\sigma_{10} = 7.21$

Figure 5.14: Response functions at different scales. Shown is the original CT scan (a) and computed responses (b) – (k). For simplicity of visualization, only one two-dimensional slice of the three-dimensional response function is shown for each of the scales. The maximum response to the nodule is achieved at the scale $\sigma_7 = 3.55$ corresponding to the size of the nodule $d = 12.33\ mm$.

Algorithm 2: Constructing nodule candidates set C

$\nabla^2_{norm} L(X, \sigma)$	computed discrete response function
S^3_X	discrete image search space
S_σ	discrete set of scales
$Lungs$	spatial lungs mask

$C \leftarrow \emptyset$	start with empty set of nodule candidates
for each $\sigma_i \in S_\sigma$	for each scale
for each $(Y, \sigma_i) \in (S^3_X \cap Lungs) \times S_\sigma$	for each discrete point in the search space
if $\nabla^2_{norm} L(Y, \sigma_i) = \max \left\{ \nabla^2_{norm} L(Z, \zeta) : (Z, \zeta) \in \mathcal{N}(Y, \sigma_i) \right\}$	if the point is search space local maxima
$C \leftarrow C \cup (Y, \sigma_i)$	update the candidates set
end if	
end for	
end for	

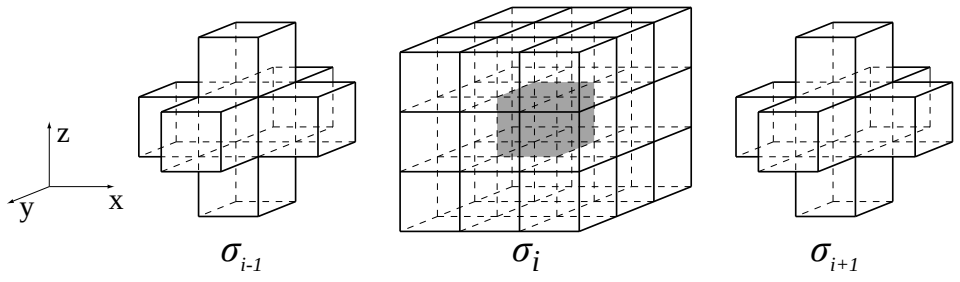


Figure 5.15: Four-dimensional neighborhood around a sample point in the discrete search space.

The third step of the algorithm finds local maxima in the four-dimensional search space (except boundary scales) with respect to both location and scale. Local maxima are determined by comparing of the filter response in a given point (Y, σ_i) to the responses of all its neighboring points $(Z, \zeta) \in \mathcal{N}(Y, \sigma_i)$. The considered neighborhood included 26 adjacent points on the discrete grid at the current scale and 7 adjacent points at scales above and below as shown in Figure 5.15.

If the filter response at the current point was maximal, i.e. it was greater or equal to the responses of all other neighboring points, the point was included to the set of nodule candidates. From the implementation standpoint, currently available computer memory (in 2010) may not be large enough to fit the entire four-dimensional search space. In this case, it is recommended to process the scales sequentially while keeping in memory only the current scale and one scale above and below.

The fourth and final step of the algorithm involves deletion of the candidates with the normalized LoG filter response value lower than the threshold $R_{CT} = 226$, which was determined earlier. An example outcome of the deletion on a selected slice of a CT image is shown in Figure 5.16. Here each circle represents a nodule candidate with diameter directly related to the scale at which it was detected.

5.2.10 Optimization of the convolution computation

One efficient implementation of the multiscale LoG transform is to do the computation of the convolutions in the frequency domain.

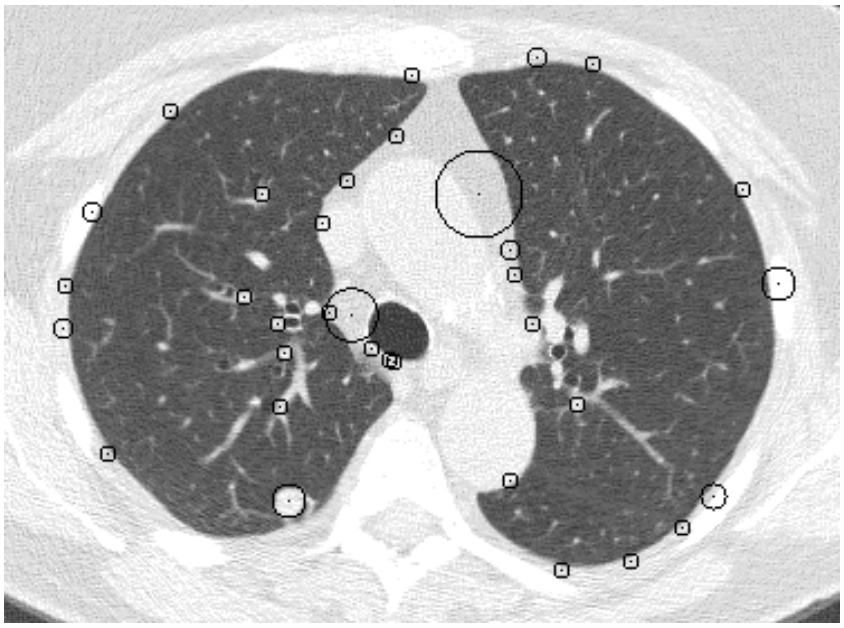
The original expression for response function for the scale σ_i

$$\nabla^2_{norm} L(X, \sigma_i) = -\sigma_i^2 \nabla^2 G(X, \sigma_i) * I(X) \tag{5.18}$$

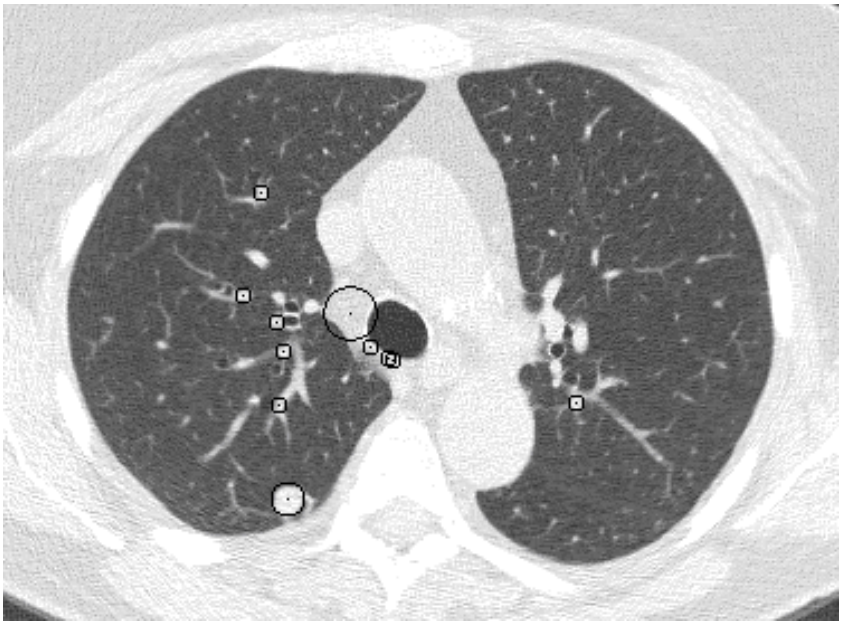
can be rewritten using the convolution theorem as

$$\begin{aligned} &\nabla^2_{norm} L(X, \sigma_i) = \\ &= -\sigma_i^2 \cdot \mathcal{F}^{-1} \left\{ \mathcal{F}\{\nabla^2 G(X, \sigma_i)\} \cdot \mathcal{F}\{I(X)\} \right\}, \end{aligned} \tag{5.19}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote forward and inverse Fourier transforms.



(a)



(b)

Figure 5.16: An output of the generator shown on one of the slices of a case: (a) original image with local maxima of response; (b) after suppression of low-response candidates.

Fourier transforms of the LoG is known and can be precomputed in advance:

$$\begin{aligned} &\mathcal{F}\left\{\nabla^2 G(X,\sigma_i)\right\} = \\ &= -(2\pi)^{-1.5}\Omega^T\Omega\exp\left(-0.5\sigma_i^2\Omega^T\Omega\right), \end{aligned} \tag{5.20}$$

where Ω is a three-dimensional vector of angular frequencies $\Omega = (\omega_x, \omega_z, \omega_z)$.

This way, in order to compute multiple convolutions for an image, one needs to take the Fourier transform once, and for each scale multiply it with precomputed transform of the LoG. Taking the inverse transform on the result and multiplying it by normalization coefficient for each scale would result in desired response function.

5.2.11 Generation of nonsolid candidates

Detection of nonsolid nodules is based on an understanding of their unique attenuation characterisitcs in a CT scan. With the exception of the image artifacts caused by heart and respiratory motion, there are no other normal volumetric structures within the lungs that have the same attenuation characteristics as nonsolid lesions. While there are a large number of voxels in the same image intensity range caused by partial volume effects at the edges of pulmonary structures, very few of them are incorporated into objects similar to nodules in size and shape. Therefore, the detection process should identify large regions of blob-like shapes having intermediate image intensity levels between parenchyma and solid tissue inside the lung region. To a first approximation, this is similar to the detection of solid nodules, but with a different target intensity range.

However, with the presence of pulmonary structures within a nonsolid re-

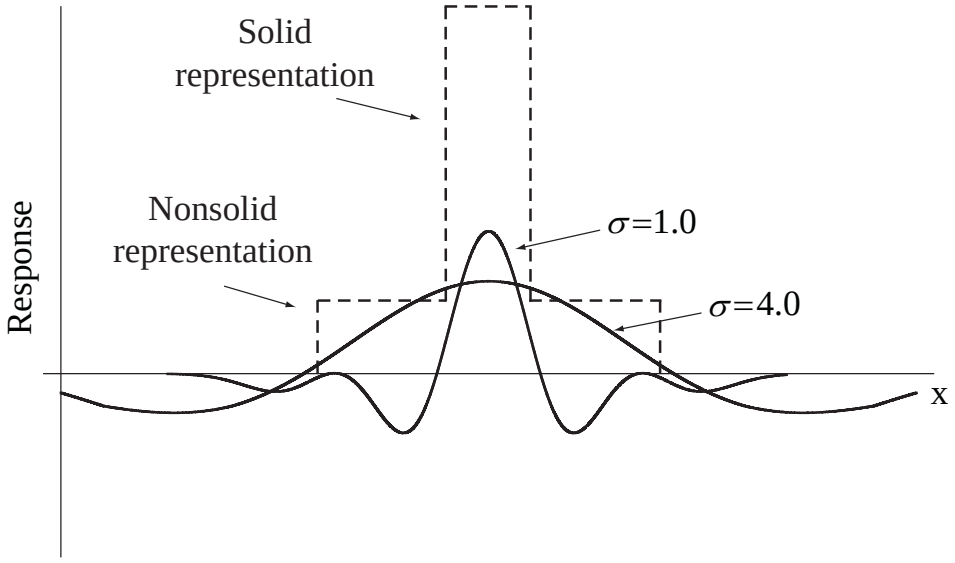


Figure 5.17: Illustration of the interference effect between solid and nonsolid nodule representations. Response to solid nodule results in higher response causing nonsolid nodule to be missed by the multiscale detector.

gion, the normalized LoG filter would produce a higher response on such structures, rather than on the entire nonsolid region due to the interference. This is illustrated in example shown Figure 5.17. Here two nodules are represented by two superimposed rectangular functions of different intensity and size. Responses of the best matching kernels are shown as solid lines. The rectangle of higher intensity resulted in higher response and therefore caused the superimposed low-intensity rectangle to be "missed" by the detector. A similar situation occurs when a nonsolid nodule is located very close to a solid pulmonary vessel.

In order to overcome this issue, the image intensity windowing was used. The main purpose of this was to suppress high intensity objects and reduce their interference. The candidate generator was modified for the detection of nonsolid nodules by preprocessing the image with a thresholding filter so that no regions of the image had a higher intensity than T. That is, the image was

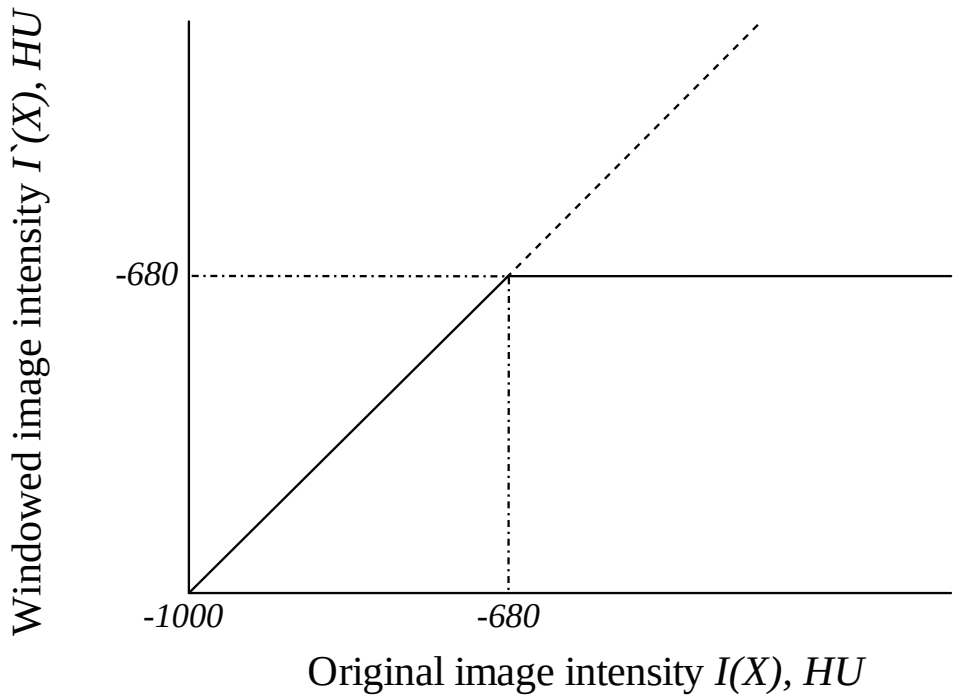


Figure 5.18: Image windowing technique. Intensity transform function is applied to the original image to suppress solid components with intensity over T.

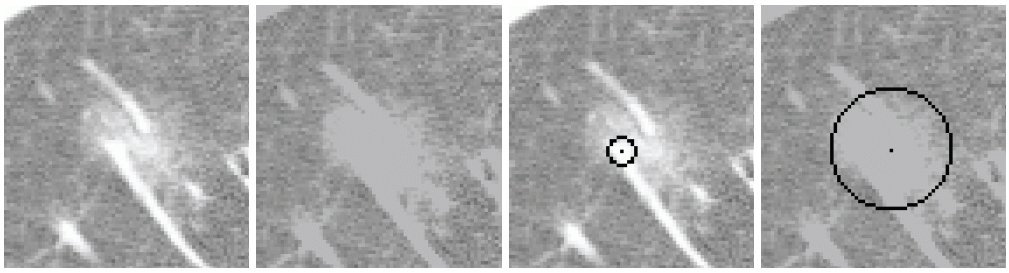
prefiltered as

$$I'(X) = \max\{I(X), T\}, \tag{5.21}$$

which is illustrated in Figure 5.18.

If we hypothesize that a solid component does not occupy more than half the volume of a nonsolid nodule, the threshold T can be set equal to the median intensity value obtained for nonsolid nodules. This way the interference from solid structures will be eliminated. The value for T was selected as -700 HU, which is 20 HU below the median intensity value obtained by sampling a set of nonsolid nodules intensities [Browder, 2007].

The illustration of such high intensity suppression on a sample nonsolid region is given in Figure 5.19. With the windowing, nonsolid regions should gen-



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

Figure 5.19: Result of windowing on a local nodule subregion at a level of -680 HU: (a) original nonsolid nodule image; (b) transformed image; (c) incorrect output of the candidate generator on the original image; (d) correct output of the candidate generator on the transformed image.

erate a strong response in LoG filtering and provide correctly located and sized nodule candidates.

Therefore, the nonsolid nodule candidate generation consists of a new prefiltering stage followed by the same nodule candidate generation used for detecting solid nodules.

5.3 Evaluation of the candidate generator

The purpose of this section is to show the practical advantage of using normalized LoG filtering scheme for identification of solid and nonsolid pulmonary nodules on a large dataset with respect to the requirements of the candidate generator. These requirements were defined in the introduction section and include: high sensitivity close, high size estimation and positional accuracy, and high speed of detection.

To show the effectiveness of the new scheme, a comparison of the normal-

ized LoG-based solid nodule candidate generator to the reference generator was made. The reference generator was previously developed within our research group and described in detail in Enquobahrie et al. [2007]. The common set of performance measures obtained from the same datasets were obtained for both of the methods.

5.3.1 Method

The candidate generation scheme was evaluated with respect to a CLDD 1.25 database. All scans in this dataset were reviewed by at least two thoracic radiologists and every identified nodule was documented. For each nodule its image location and measured length and width were recorded. The effective diameter $d(k_i)$ of the nodule k_i from its width $w(k_i)$ and height $h(k_i)$ were determined by:

$$d(k_i) = \frac{1}{2}\left(w(k_i) + h(k_i)\right). \tag{5.22}$$

The evaluation dataset consisted initially of 250 sequential asymptomatic cases from a lung cancer screening study [Henschke et al., 1999], and was enriched with 456 new cases containing at least either one solid nodule with effective diameter greater or equal to 4 mm or one nonsolid nodule greater or equal to 6 mm. With the addition of the enriched data, the fraction of nodules greater than 10 mm increased from 3% to 6% and the fraction of nodules greater than 4 mm increased from 23% to 43%.

Nodules with effective diameter greater or equal to the lower size cut off threshold made up the evaluation set. In the experiment, solid nodules of an effective measured diameter equal or greater than 4 mm (499 nodules) were

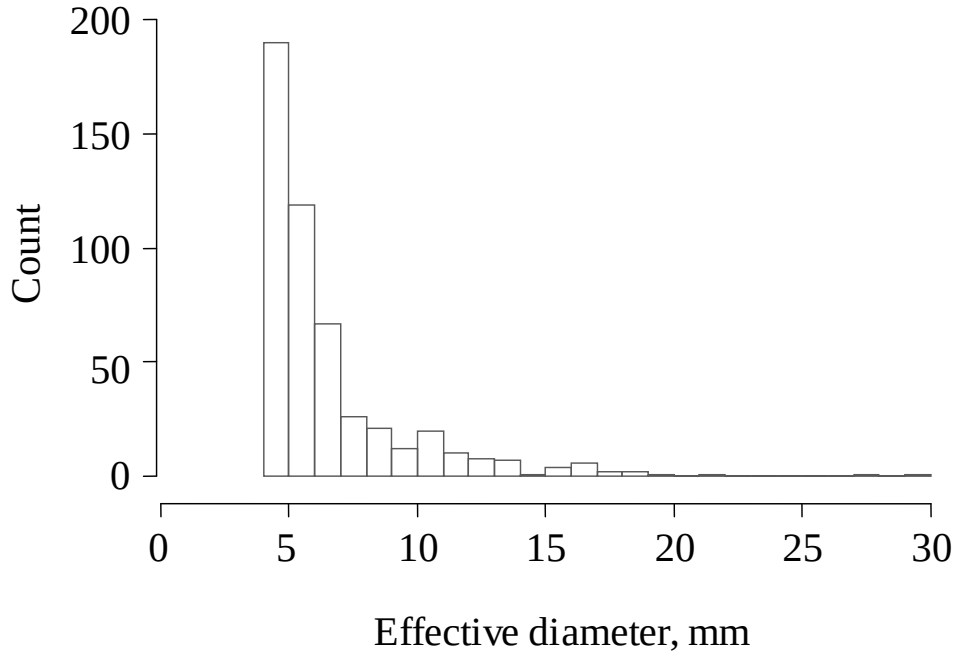
used. For nonsolid nodules, the diameter cut off was selected to be 6 mm (107 nodules). The value of both size thresholds was slightly lower than the one used in clinical practice to establish the safety margin covering the disagreement between automatic and manual nodule size measurements. The distributions of the nodule sizes for the final dataset are shown in Figure 5.20.

The conventional way to assess sensitivity is to measure the fraction of correctly identified nodules over the total number of nodules in the evaluation set. If we denote the set of all nodules that are being targetted as K, the final sensitivity S can be calculated as the fraction of correctly identified nodules n_c over the total number of nodules $|K|$. The criterion, confirming that a nodule has been detected, is important and may have an impact on the sensitivity: a nodule k_i was considered as identified, if there existed a corresponding candidate c_j, such that the Euclidean distance between their centroid locations was less or equal to the half-length of the nodule k_i. The Algorithm 3 shows how the sensitivity is calculated.

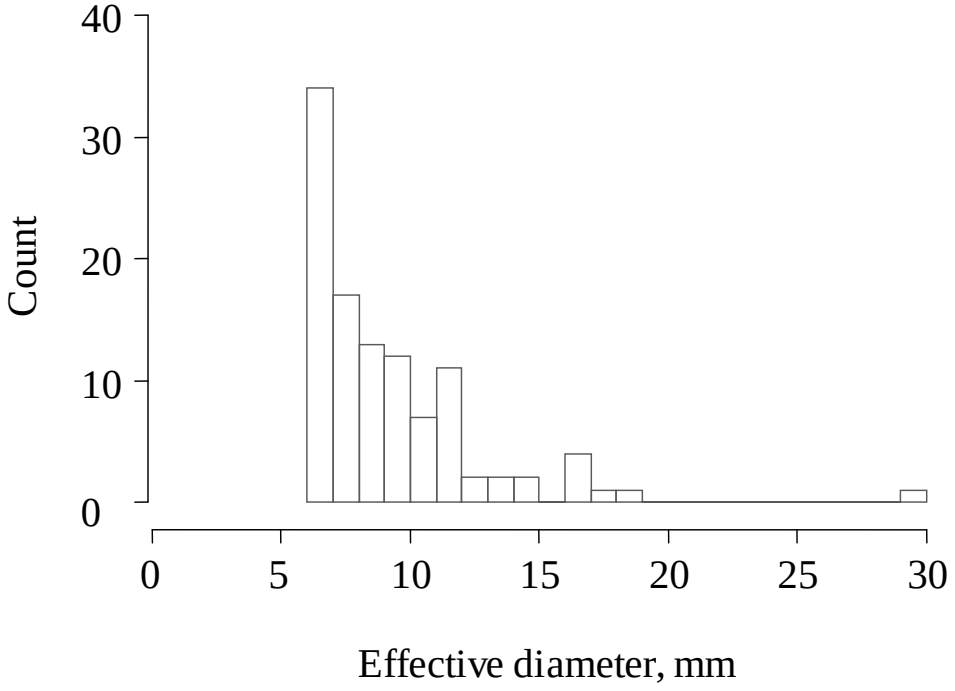
The sensitivity was evaluated independently for solid and nonsolid nodules with the intensity windowing preprocessing.

The accuracy of the size estimation was evaluated through a comparison of the effective nodule diameter derived from the ground truth to the diameter obtained from the generator. The bias for diameter estimation for the set of detected nodules was calculated as the average difference between the effective diameter $d(k_i)$ of the nodule and the diameter of corresponding candidate as $d(c_j)$:

$$DiameterBias = \underset{c_j \leftrightarrow k_j}{E} \left[d(k_i) - d(c_j) \right]. \tag{5.23}$$



(a)



(b)

Figure 5.20: Distributions of solid (a) and nonsolid (b) nodule effective diameters in the evaluation dataset.

Algorithm 3: Calculation of detection sensitivity

K	evaluation set of nodules
C	set of detected candidates

$n_c \leftarrow 0$	set the number of correctly identified nodules
for each $k_i \in K$	for each nodule
for each $c_j \in C$	for each candidate
if $dist(c_j, k_i) \leq 0.5\, l(k_i)$	if the distance between the centroids is less than the half length of the nodule
$n_c \leftarrow n_c + 1$	increment the number of correctly identified nodules
end if	
end for	
end for	
$S \leftarrow n_c/\|K\|$	calculate the sensitivity

Positional accuracy was evaluated by calculating the average distance between the centroids of the detected nodules and corresponding candidates.

Average candidate generation execution time per CT image was measured by running the candidate generators on Intel Xeon 3.00 GHz processor given that the $Lungs$ mask is already segmented.

5.4 Results

The sensitivity for solid nodules achieved a value of 0.998 (498/499). One solid nodule in the apical region of the lungs was missed due to the coarseness of the lung segmentation.

The normalized LoG-based candidate generator considered in this thesis

Table 5.3: Comparison of LoG-based and benchmark generators with respect to detection of solid nodules with of diameter of 4 mm and above.

Parameter	LoG-based	Reference
Sensitivity	0.998 (498/499)	0.958 (478/499)
Diameter estimation accuracy, mm	-0.12 $\pm$ 3.27	-1.20 $\pm$ 5.45
Positional accuracy, mm	1.40	1.66
Size range of detected candidates, mm	3.00 – 25.00	1.96 – 22.34
Average number of candidates per case	8177	3785
Average CT image processing time, minutes	4.5	8.0

outperformed the reference generator in identifying solid nodules. The comparison of the generators with respect to the solid nodules of diameters 4 mm and greater is shown in Table 5.3.

Sensitivity of the reference generator on the evaluation set was lower. Its scheme for detection of attached nodules was highly dependent on the correctness of the lung segmentation, therefore, the majority of the false negatives were located on periphery of the lung. Even though both LoG-based and reference generators resulted in an overestimation bias for nodule diameter, the LoG-based generator was substantially more accurate and resulted in smaller error: the 95% confidence intervals of the difference between automated and manual nodule size measurements were -0.12 $\pm$ 3.27 mm for solid nodules.

The normalized LoG-based candidate generator also turned out to be more accurate than the reference generator in centroid estimation as its average nodule candidate distance of 1.40 mm was lower than 1.66 mm. This could be partially explained by the fact that the reference algorithm locates the peaks for

attached nodules as opposed to their centroids.

For nonsolid nodules, the LoG-based candidate generator achieved the perfect sensitivity value of 1.000 (107/107).

The generator of nonsolid candidates resulted in a size overestimation of -1.27 $\pm$ 5.70 mm, while the average candidate location accuracy was 1.43 mm for nonsolid nodules.

Each solid and nonsolid candidate generation algorithms were able to complete the processing of a single scan in 4.5 minutes on average.

5.5 Discussion

Both solid and nonsolid candidate generators achieved very high detection sensitivities of 99.8% and 100.0%, which makes the detection systems based on the normalized LoG filter a promising solution for the detection of pulmonary nodules. Candidate generation with windowing prefiltering has been found to be a powerful technique.

LoG-based candidate generation resulted in size overestimates for both solid and nonsolid nodules. One of the possible reasons for this is that the multiscale LoG response function is asymmetric (5.7). The overestimate was greater for nonsolid nodules, probably, because of greater average size, and the subjectivity in size measurements by radiologists: the boundary of a nonsolid nodule is not clearly defined and therefore the bias in candidate size estimation may also depend on how the operator set up the windowing level.

The experiments have also shown that nonsolid nodules could be identified with almost the same positional accuracy (1.40 mm vs. 1.43 mm) as solid nodules.

The distribution of solid candidate responses for the set prior to deleting candidates with low filter response is shown in Figure 5.6. The graph has three local modes that were visually identified as corresponding to: (a) small intensity variations (noise) in lung parenchyma; (b) lung and mediastinal surface irregularities including some attached nodules; (c) pulmonary vessels, airways, tips of the ribs, calcification and remaining nodules. The lowest normalized response for a solid nodule was 228, as compared to the theoretically computed value of 226 which means all the true candidates above the threshold were preserved. The result of rejecting low-response candidates is illustrated in Figure 5.22. The plot shows the distribution of candidate sizes with respect to their estimated size. From this graph one can infer the expected number of candidates generated for a given size range. The number of candidates increases rapidly with the decrease of candidate size and therefore, performance of the generator can be fine-tuned depending on what the target nodule size range is.

The speed of the presented candidate generator is mostly limited by the speed of calculating discrete Fourier transforms and is inversely proportional to the selected number of scale discretization levels. Further optimization of the convolution computation can be done by employing more efficient filtering algorithms, such as suggested by Jin and Gao [1997].

The proposed candidate generator achieves higher sensitivity but at a cost of a significant increase in false positives (this is also the case for our reference method) when compared to other candidate generators reported in the litera-

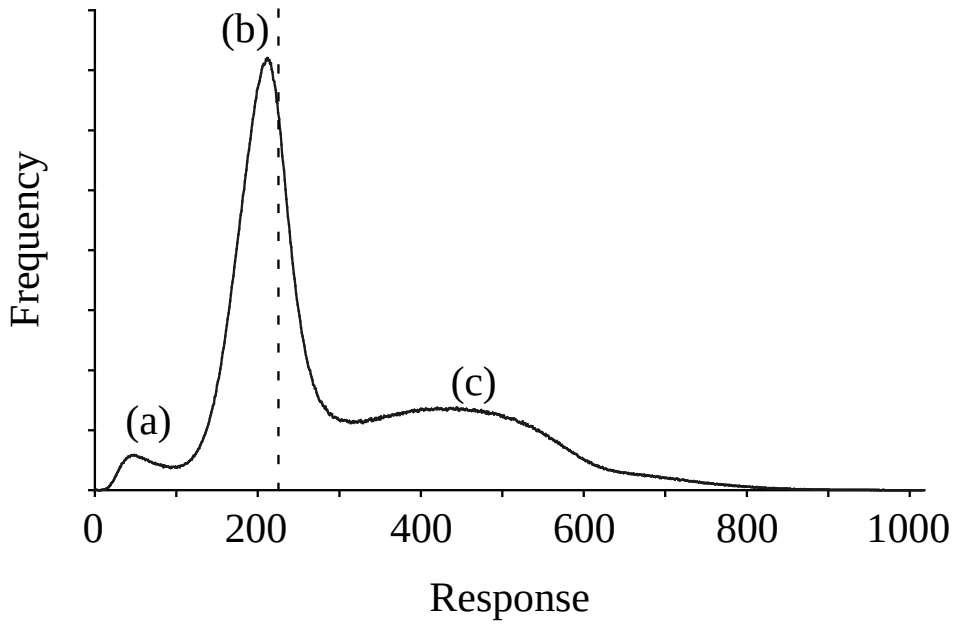


Figure 5.21: Distribution of the candidate normalized responses shows three distinct modes corresponding to: (a) noise in lung parenchyma, (b) lung surface irregularities including attached nodules and (c) pulmonary vessels, airways and remaining nodules with lesser degree of attachments. Dashed line is the response threshold.

ture (Table 5.1). However, elimination of the vast majority of those false positives is possible with very simple filters that impose shape and density constraints on the candidates [Enquobahrie et al., 2007]. Therefore this added burden of additional false positives is not a major impediment to implementation of a full CAD system while the improvement in sensitivity of the candidate generator will have a major impact on the overall performance of the CAD system.

5.6 Conclusion

In this chapter, a method based on multiscale LoG filtering for initial candidate generation in the task of automated detection of pulmonary nodules was pre-

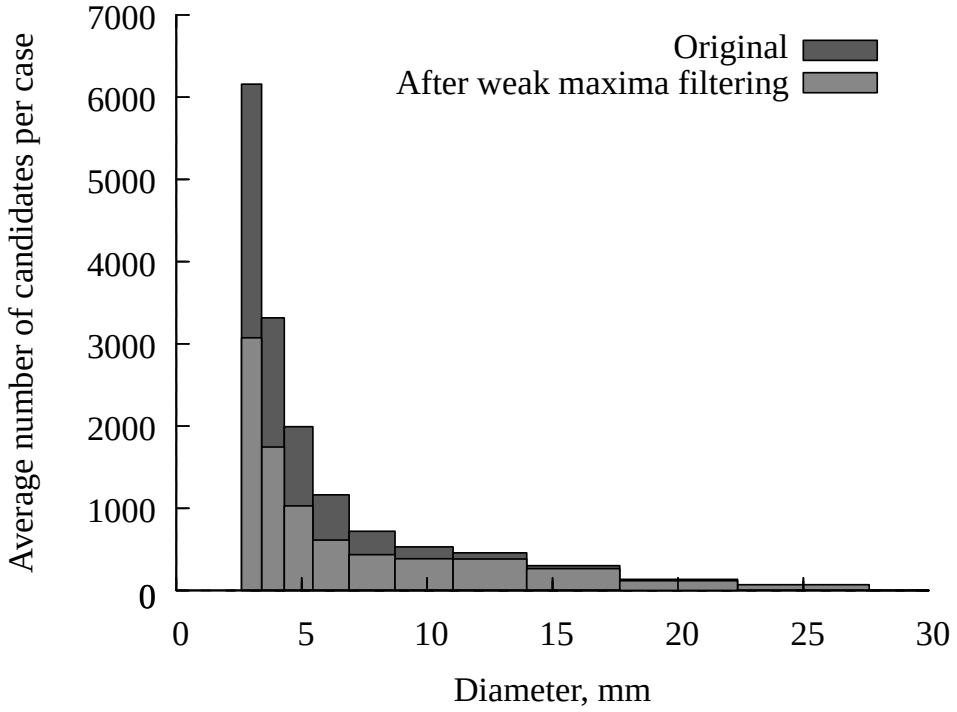


Figure 5.22: Distribution of candidate sizes before and after suppression of low response candidates. Each bin corresponds to one of 10 discrete diameter levels used by the generator.

sented and evaluated. The sensitivity of the candidate generator was shown to reach 0.998 (498/499) for solid nodules. The sensitivity of the generator for nonsolid nodules achieved 1.000 (107/107) after the application of the intensity windowing technique. Further, application of scale-normalized LoG filters allowed for both accurate size and nodule centroid estimation. These results suggest that multiscale LoG filtering is a very effective tool for the detection of pulmonary nodules from whole-lung CT scans.

CHAPTER 6

STANDARD MOMENTS-BASED VESSEL BIFURCATION FILTER

Achieving both high sensitivity and low false positive rate is the key purpose of an automated nodule detection system. While isolated intraparenchymal nodules with no vascular attachments, or very few, are easy to detect, lesions with significant attachments are often confused with blood vessels of different morphological variations. If the detection system is configured for high sensitivity, pulmonary vessel junction points become a significant source of false positive findings [Agam et al., 2005; Enquobahrie et al., 2007; Das et al., 2006; Pu et al., 2008; Lee et al., 2004].

There are several types of false positives caused by vessel junction points. The most common junction is a vessel bifurcation where a parent branch splits into two children branches of approximately equal and smaller diameters as illustrated in Figure 6.1(b). The morphology of such a junction is somewhat similar to a nodule attached to a vessel of comparable size shown in Figure 6.1(a) and, therefore, poses a challenge for an automated detection method. A second possible case is rarer and is shown in Figure 6.2: one child branch could be much smaller while another one has roughly the same diameter and is collinear with the parent branch. The remaining large class of vascular structures contains more complex junctions and crossings of multiple branches. An example is given in Figure 6.3.

This particular work is focused on a method for the discrimination of true nodules from the most common class of vessel bifurcations. Even though the morphology of the vessel bifurcation is relatively straightforward, it is difficult to devise a simple method that will robustly distinguish them from pul-

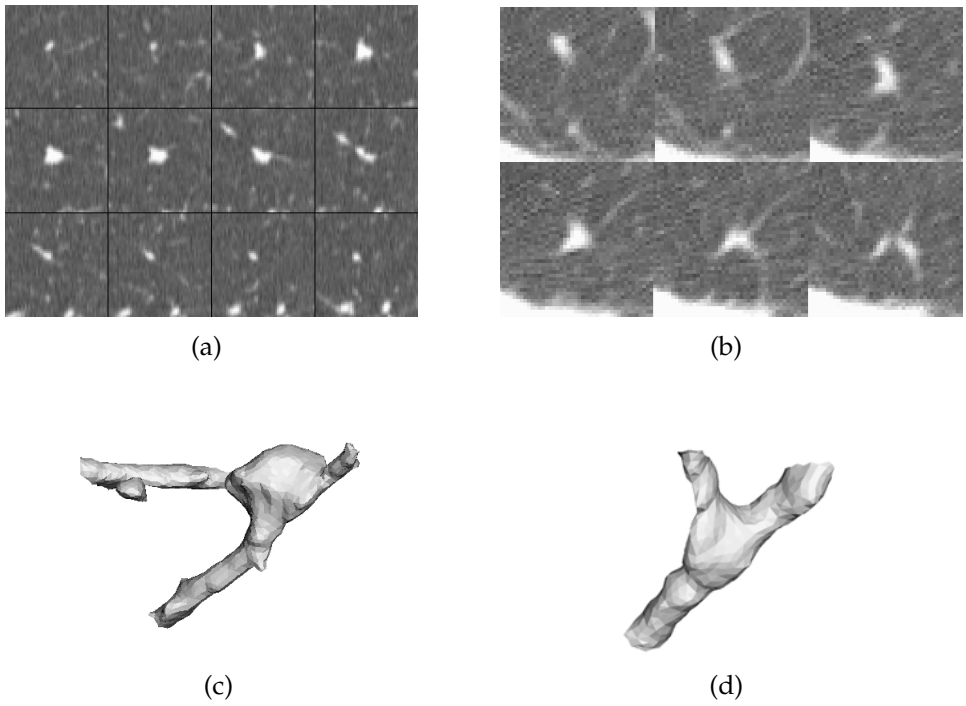


Figure 6.1: An example of typical candidate shapes to be discriminated: vascularized nodule (a) and pulmonary vessel bifurcation point (b). Light shaded 3D rendering after thresholding the images at the level of -400 HU are shown in subfigures (c) and (d).

monary nodules. Moreover, this type of junction point often occurs near the mediastinum region where the pulmonary vessels are affected by heart motion or pass close to the airway tree, which complicates the discrimination. One hypothesis is that a series of methods that target a specific class of false positives will yield a better performance than the commonly used universal filters that target all possible false positive types at the same time.

While there has been much research on the discrimination of a broad class of false positives from nodules, surprisingly little attention has been devoted to the particular classes of false positives and vessel bifurcations in particular. Zhao et al. [Zhao et al.] proposed a parametric model of the bifurcation point made of

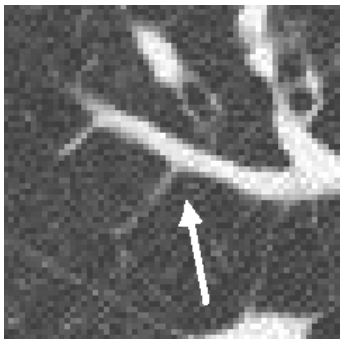
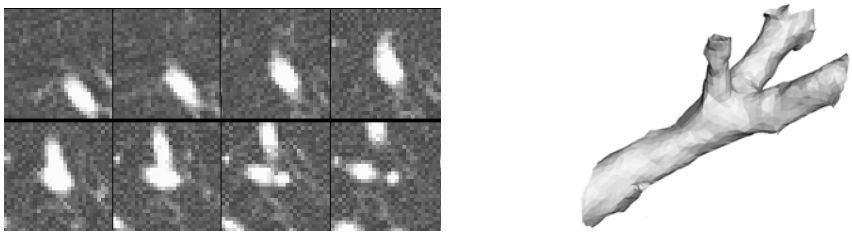


Figure 6.2: An example of a vessel bifurcation, showing unequally sized child branches.



(a) (b)

Figure 6.3: An example of complex pulmonary vessel junction point: (a) montage view and (b) 3D rendering.

three toroidal components. Local principal curvatures were used to map a pixel to either a junction, a vessel, or an ellipsoid nodule model, with subsequent classification. Bahlmann et al. [Bahlmann et al., 2006] used Gaussian model fitting followed by extraction of a manifold containing information on the bounding sphere and further analysis. Even though the technique looks promising, it has not been evaluated quantitatively. Several attempts were made to extract the entire pulmonary vasculature tree from a CT scan with the purpose of eliminating false positives caused by vessel junctions. The works of Croisille et al. [Croisille et al., 1995] and Agam et al. [Agam et al., 2005] showed some improvement in false positive reduction; however, in addition to the high complexity of the method, the sensitivity of the detection suffered as well. Earlier work of our

research group [Enquobahrie et al., 2007] employed a feature that assesses the extent of an attachment relative to a size of a candidate: a vessel junction point would have less enclosed volume than a true nodule for the same amount of attachment. Another technique that was used for the discrimination of vessel junctions is the analysis of the candidate's principal axes computed from moments. True round-shape nodules would have high compactness, or a low ratio of the largest to smallest axis of the ellipsoid of inertia. In contrast, the smallest axis of the bifurcation point region would be considerably smaller than the largest axis. All mentioned techniques had only limited success in the past.

6.1 Method for discriminating bifurcation points

The method presented here does not seek to construct an explicit geometric model of a nodule or a bifurcation point, but rather to capture the difference between these two shapes from their geometric signatures. Since a round nodule and a vessel bifurcation are two fundamentally different objects, they must have different rotational properties and, consequently, distinct sets of moments. Size and orientation of the same class of objects might be different within the class; therefore normalization is needed, which is achieved by the method of three-dimensional standard moments. Standard moments provide a numerical characterization of an object shape that can be used to discriminate between true nodules and pulmonary vessel junctions.

A multiscale Laplacian of Gaussian (LoG) based detector identified nodule candidates from a whole-lung CT scan and supplied their centroids and radii as discussed in Chapter 5. Selection of a specific candidate generator is not important for the presented method as long as it provides the approximate location of

nodule candidates.

In this thesis, geometric moments (based on binary segmented image) were chosen over density moments (computed from the raw pixel values), since shape carries more information than density does in the discrimination of same intensity solid objects. Geometric moments are different from density-based moments by the fact that they are computed for binary images, where the intensity values are set either to one (foreground) or zero (background). Once the moments are computed, they can be used to characterize the candidate shape.

The full process for discriminating pulmonary vessel bifurcations from nodules for a given candidate location and size is as follows:

1. Candidate subimage preprocessing.
   (a) resampling to isotropic space
   (b) intensity thresholding
   (c) bounding the candidate subregion with a sized sphere
2. Calculation of the standard moments set (normalized for size and rotation).
3. Classification of the moments vector.

These steps are discussed in detail later in this chapter.

6.1.1 Candidate preprocessing

The preprocessing step starts with isotropic resampling of the candidate subregion to equalize the image resolution along the three coordinate axes.

The binary image $B(x, y, z)$ of the candidate was obtained by thresholding the candidate subregion of the isotropic image $I(x, y, z)$ using value of T:

$$B(x,y,z) = \begin{cases} 1, & \text{if } I(x,y,z) > T; \\ 0, & \text{otherwise.} \end{cases} \tag{6.1}$$

This threshold must separate the solid tissue from the surrounding parenchyma and provide the foreground region for the subsequent raw moments calculation.

In order to compute the set of raw moments, the region of interest $t(x, y, z)$ must first be selected to incorporate as much useful shape discrimination information as possible. This is achieved by imposing a spherical mask $S(R)$ of radius R centered over the candidate and rejecting all image data of the candidate subregion that is outside the mask:

$$t(x,y,z) = \begin{cases} B(x,y,z), & \text{if } (x,y,z) \in S(R); \\ 0, & \text{otherwise.} \end{cases} \tag{6.2}$$

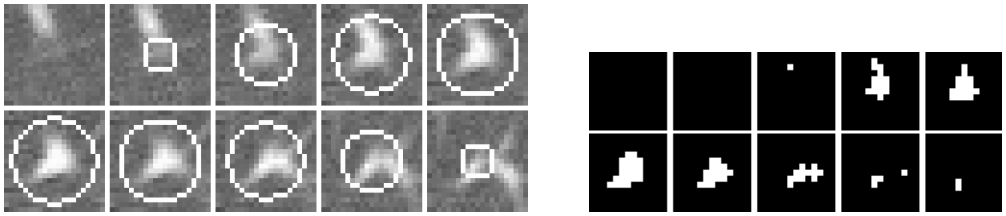
Here the selection of radius R is important: if the radius is small relative to the extent of the candidate, the window will not contain any lung parenchyma and there will be no way of discriminating the shapes. Conversely, if the window is too large, it may contain other pulmonary structures that will negatively affect the calculation and introduce unnecessary noise. This radius should depend on the size of the candidate, but in the general case, such information is not provided. To address this issue, the concept of a foreground fraction $F(R)$ is introduced. This fraction is equal to the ratio of foreground volume to the total volume of a spherical mask $S(R)$.

$$F(R) = \frac{\sum\limits_{x,y,z \in S(R)} B(x,y,z)}{\sum\limits_{x,y,z \in S(R)} 1}. \tag{6.3}$$

The radius R of the spherical mask is the smallest one among those resulting in foreground fraction no greater than parameter P:

$$R = min\Big\{r : F(r) \leq P\Big\}, \tag{6.4}$$

or, in other words, the spherical mask is selected such that the fraction of the foreground component inside approximately equals P, the same for all candidates. An example of the initial candidate subregion and the results of the preprocessing are shown in Figure 6.4. The optimal values for the parameters T and P were determined by experiment.



(a) (b)

Figure 6.4: An example of original candidate subregion showing the surface of the masking sphere (a); result of intensity thresholding and subregion masking (b).

6.1.2 Calculation of the standard moments set

The method of standard moments was first applied to the classification of three-dimensional shapes by Reeves and Wittner [Reeves and Wittner, 1983]. Standard moments have the advantage over raw moments because they are normalized with respect to translation, rotation, and scale transformation and remain invariant for a given object. They can be used to detect and characterize objects that may be located anywhere within a space of interest, may be arbitrarily oriented and may have different size.

Raw three-dimensional moments of order $p+q+r$ for the candidate region of interest $t(x,y,z)$ are defined by:

$$M_{pqr} = \sum_{x,y,z \in S(R)} x^p y^q z^r t(x,y,z). \tag{6.5}$$

Standard moments S_{pqr} can be directly computed from raw moments by normalization. First, the volume of the object is scaled to be 1:

$$M'_{pqr} = \lambda^{3+p+q+r} M_{pqr},\ \lambda = (M_{000})^{-\frac{1}{3}}. \tag{6.6}$$

Translation normalization is achieved by shifting the origin of the coordinate system to candidate's center of mass by the following transformation:

$$M''_{pqr} = \sum_{s=0}^{p}\sum_{t=0}^{q}\sum_{u=0}^{r} \binom{p}{s}\binom{q}{t}\binom{r}{u} a^{p-s} b^{q-t} c^{r-u} M'_{stu}, \tag{6.7}$$

where $a = -M'_{100}$, $b = -M'_{010}$, $c = -M'_{001}$.

Rotation normalization is done by aligning the candidate's principal axes with the coordinate axes:

$$\begin{aligned} S_{pqr} = M'''_{pqr} = \sum_{s_1=0}^{p}\sum_{t_1=0}^{s_1}\sum_{s_2=0}^{q}\sum_{t_2=0}^{s_2}\sum_{s_3=0}^{r}\sum_{t_3=0}^{s_3} \binom{p}{s_1}\binom{s_1}{t_1}\binom{q}{s_2}\binom{s_2}{t_2}\binom{r}{s_3}\binom{s_3}{t_3} \cdot \\ u_{11}^{t_1} u_{21}^{t_2} u_{12}^{s_1-t_1} u_{22}^{s_2-t_2} u_{13}^{p-s_1} u_{33}^{q-s_2} u_{31}^{t_3} u_{32}^{s_3-t_3} u_{33}^{r-s_3} \cdot \\ M''_{t_1+t_2+t_3, s_1+s_2+s_3-t_1-t_2-t_3, p+q+r-s_1-s_2-s_3}, \end{aligned} \tag{6.8}$$

where $U = (u_{ij}) = (\pm u_1, u_2, u_3)^T$ is made of the orthonormal set of eigenvectors of the matrix N, which is, in turn, composed of the second order scaled and translated moments:

$$N = \begin{pmatrix} M''_{200} & M''_{110} & M''_{101} \\ M''_{110} & M''_{020} & M''_{011} \\ M''_{101} & M''_{011} & M''_{002} \end{pmatrix}. \tag{6.9}$$

The sign of u_1 is selected such that $det(U) = 1$. If $\{u_i\}$ are sorted in the order of descending eigenvalues, than the object's largest principal axis will be aligned along x and the smallest one along z coordinate axes without ambiguity. The standard set of moments up to order three $p + q + r \leq 3$ is computed. Some of the moment values are trivial and disregarded — the volume of the object after normalization is equal to 1: $S_{000} = 1$; the center of mass is at the origin of the coordinate system: $S_{100} = S_{010} = S_{001} = 0$; and the principal axes of the ellipsoid of inertia lie on the coordinate axes: $S_{110} = S_{101} = S_{011} = 0$. Finally, the following set of remaining thirteen standard moment values is used in classification: S_{200}, S_{020}, S_{002}, S_{300}, S_{030}, S_{003}, S_{201}, S_{210}, S_{120}, S_{102}, S_{111}, S_{021}, S_{012}.

6.1.3 Classification of the moments vector

In the context of the high dimensionality of the feature vector, where the importance of each individual component is unknown, a soft margin support vector machine (SVM) with a polynomial kernel was chosen for classification. The SVM^{light} package [Joachims, 1999] was used in the implementation.

6.2 Experimental setup and data

The evaluation set consisted of 276 intraparenchymal nodules with a diameter no less than 4 mm from the documented dataset of 656 low-dose whole-lung CT scans with a slice thickness of 1.25 mm obtained from Weill Cornell Medical Center. This dataset is maintained by our research group and was extensively used in the previous studies [Enquobahrie et al., 2007]. In our experimental

setup the cases with an even case identifier were used for training and optimization, while the odd-numbered cases were used only for final testing. Accordingly, 130 nodules were used for training and 146 nodules for testing. In addition to true nodules, the same number of bifurcation points was manually sampled from the set of available candidates provided by our nodule candidate generator.

The discrimination scheme was optimized exclusively on the training set using five-fold cross-validation. The parameters to optimize were foreground volume fraction P and intensity threshold T. A false positive reduction filter should preserve all true nodules and filter out false candidates; therefore we used the false positive reduction fraction obtained at 100% sensitivity to true nodules as the performance measure. For example, if a certain configuration of the filter reduced the number of vessel bifurcations from the original 130 to 65 without the loss of sensitivity to true nodules, the corresponding reduction fraction would be equal to 0.5. The optimization was done jointly for the foreground volume fraction P and the threshold T. The configuration, incorporating the optimal parameters and classification model, was applied to the test set and the resultant false positive reduction was obtained.

Once the optimal values for P and T were determined, the performance of the resulting classifier was compared to previously developed techniques of (a) principal curvatures [Koenderink and van Doorn, 1992; Li et al., 2003], (b) ellipsoid of inertia compactness, and (c) attachment ratio [Enquobahrie et al., 2007]. The area under the Receiver Operating Characteristic (ROC) curve (AUC), obtained for the task of discrimination between nodules and vessel bifurcation points on the test set, was used as the performance metric.

Finally, the significance of the standard moments based filter was evaluated on the experimental nodule detection system previously developed by our research group. The free-response receiver operating characteristic (FROC) curves were generated for the test set before and after application of the filter. In addition, false positive reduction fractions were calculated at different levels of detection sensitivity.

Table 6.1: The values of false positive reduction fraction obtained for different foreground volume fractions and intensity thresholds on the training set.

		Foreground volume fraction P				
		0.1	0.2	0.3	0.4	0.5
Threshold T, HU	-250	0.43	0.61	0.08	0.06	0.01
	-300	0.47	0.68	0.30	0.11	0.02
	-350	0.36	0.71	0.54	0.20	0.13
	-400	0.29	0.69	**0.75**	0.28	0.17
	-450	0.28	0.64	0.63	0.28	0.18
	-500	0.27	0.61	0.51	0.32	0.22
	-550	0.27	0.43	0.42	0.37	0.27
	-600	0.17	0.42	0.42	0.31	0.18
	-650	0.16	0.34	0.39	0.30	0.15

6.3 Results and discussion

The results of the optimization on a training set with respect to the false positive reduction fraction are shown in Table 6.1. The values $P = 0.3$ and $T = -400HU$

that corresponded to the reduction fraction of 0.75 at 100% sensitivity were selected as optimal and applied to the test set. The resultant false positive reduction fraction on the test set was 0.80 at a sensitivity of 99%: one nodule out of 146 was rejected as a result of the filtering.

The resulting ROC of the model with optimal parameters is given in Figure 6.5 together with the classification performance of the other methods applied individually to the same test dataset. The standard moments method resulted in a higher AUC of 0.99 compared to the conventional shape analysis techniques based on principal curvatures (AUC=0.89) and the ellipsoid of inertia (AUC=0.86). Moreover, the method provides better discrimination between nodules and bifurcation points than the attachment ratio feature (AUC=0.97), previously employed by our detection system. ROC plot indicates that at very high sensitivity settings close to 100%, the standard moments based filter had the highest specificity among considered methods.

A comparison of the FROC curves for the experimental detection system, obtained with and without standard moments based filter, is given in Figure 6.6. An alternative way to represent the change in the detection performance is shown in Figure 6.7. Here the values of false positive reduction fraction are given at different levels of detection sensitivity. These graphs show that the standard moment method results in a significant reduction of the false positive rate especially at high detection sensitivity levels, where the pulmonary vessels are still very often confused with true nodules. However, at low sensitivity levels, the filters, previously used in the experimental detection system, are capable of rejecting vessel bifurcations. Consequently, the benefits of using this vessel bifurcation filter are very low for these sensitivity rates.

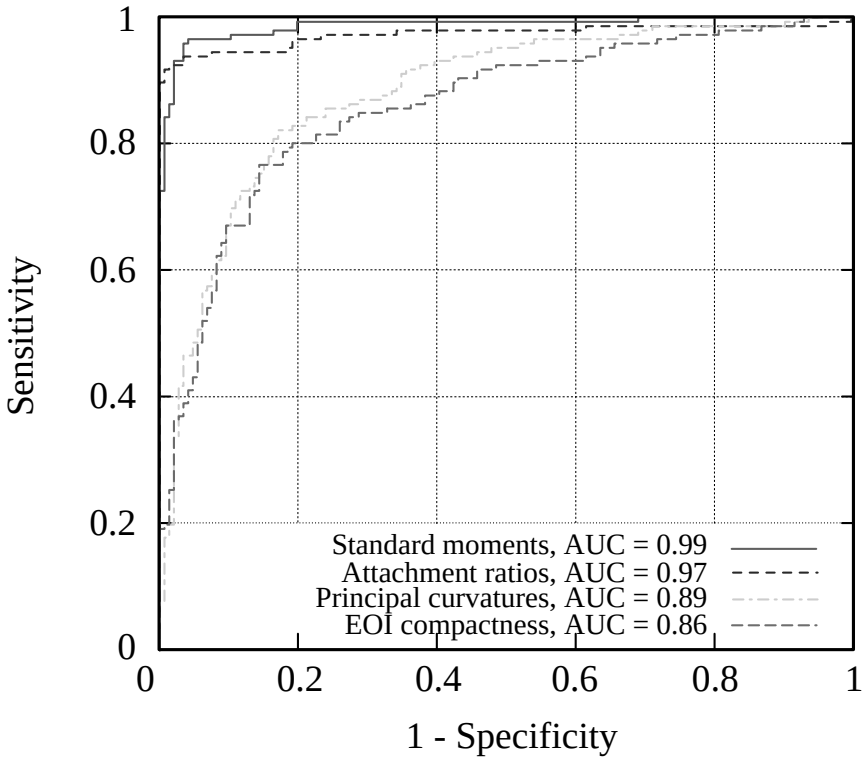


Figure 6.5: Performance of individual predictors for nodule – vessel bifurcation point discrimination on the test set.

An example where the standard moment method achieves an advantage over previously used techniques is shown in Figure 6.8. Partial volume effect "erodes" the thin branches of a small bifurcation point, while the junction point remains. It makes the use of attachment and ellipsoid of inertia filters inefficient. However, a new filter was able to distinguish the geometry of branches and classify the example properly.

Example cases where the nodules and bifurcation points were misclassified by the moment filter are shown in Figure 6.9. Both cases have a geometric morphology different from typical representatives of the class: the bifurcation point in Figure 6.9(a) is adjacent to other pulmonary vessels, while the nodule in Fig-

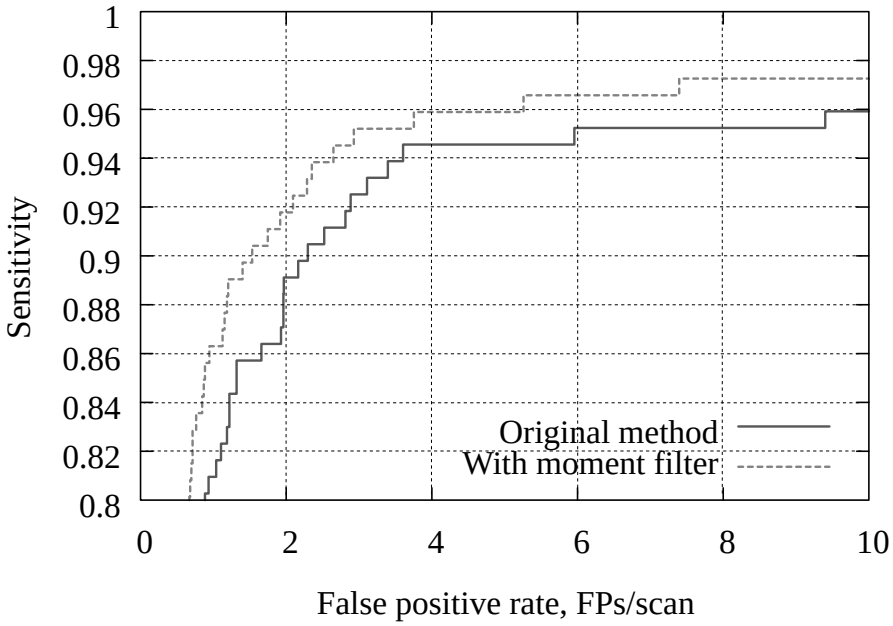


Figure 6.6: Performance of the moment-based filter in experimental detection system: FROC curves obtained with and without the standard moment based filter.

ure 6.9(b) has an unusual elongated shape with multiple vessel attachments.

6.4 Conclusion

A standard moments based vessel bifurcation filter for computer-aided detection of pulmonary nodules is presented and evaluated. The method resulted in 99% sensitivity and 80% specificity for the task of distinguishing nodules from vessel bifurcation points. Moreover, the filter was able to reject up to 69% of the false positives with almost no loss in sensitivity after incorporating it into design of an experimental detection system. These results suggest that the method of standard moments is very effective for rejecting bifurcation false positives for

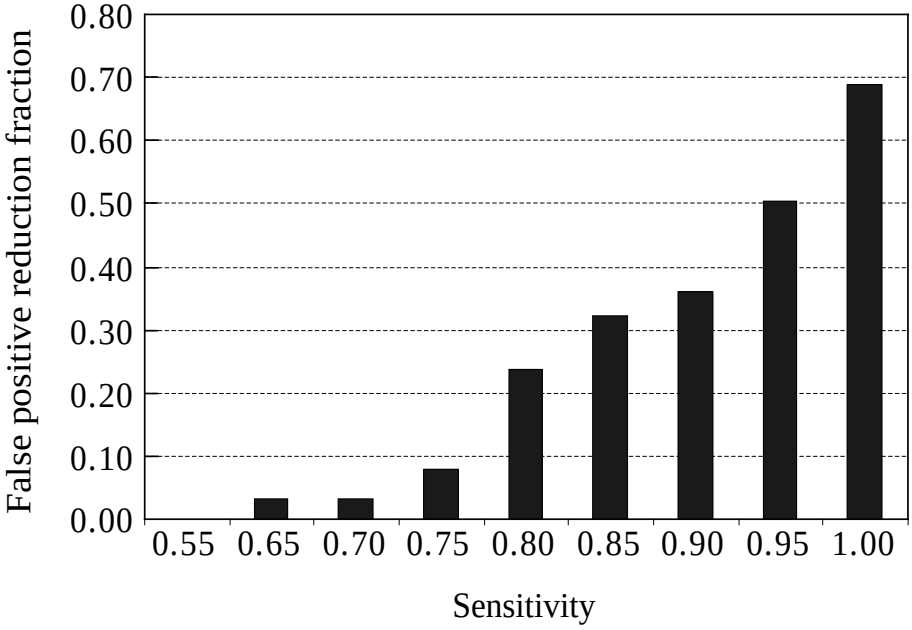
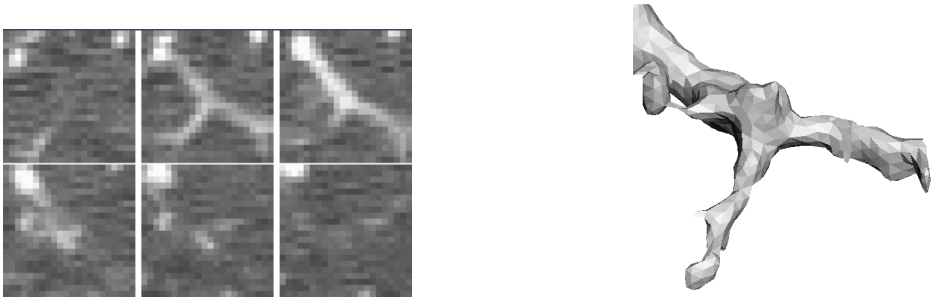


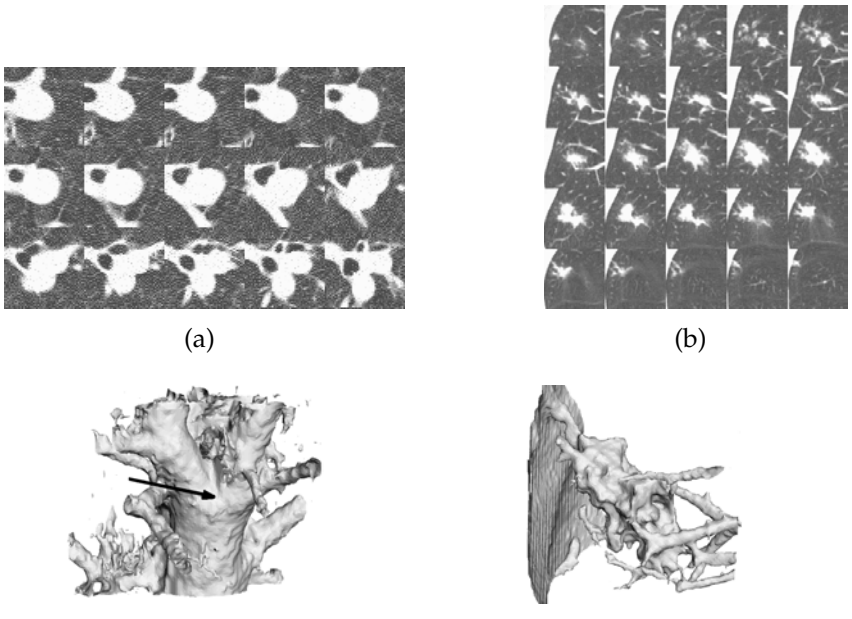
Figure 6.7: Performance of the moment-based filter in experimental detection system: false positive reduction fraction at different levels of sensitivity.

the task of automated nodule detection.



(a) (b)

Figure 6.8: An example of small bifurcation protruded due to partial volume effect. While it was correctly identified by the moment filter, it was confused with an attached nodule by both ellipsoid of inertia and attachment filters: montage view (a) and 3D rendering (b).



(a) (b)

(c) (d)

Figure 6.9: An example of incorrectly classified candidates: an axially oriented bifurcation point with multiple side attachments was identified as a nodule (a); a nodule of irregular shape was confused with a bifurcation point (b). Corresponding 3D visualizations are shown in subfigures (c) and (d).

CHAPTER 7

QUANTIFICATION OF NODULE SIZE MEASUREMENT ERROR

In the domain of computerized detection, the trade-off between sensitivity and number of raised false positives is an important indicator of an algorithm's performance [Bowyer, 2000; Chakraborty and Winter, 1990]. Both metrics are derived from a comparison of the expert decision, the ground truth, to the decision of the algorithm. The aim for the majority of automated detection systems is to approach the ground truth: to maximize sensitivity and to minimize the number of false positives. However, in nodule detection, nodule size is an important factor that affects the ground truth and complicates performance evaluation and comparison of various automated systems.

For any documented set of scans of pulmonary nodules, there is a minimum size limit due to image noise and reconstruction resolution below which nodules cannot be reliably seen. Very small nodules may be barely distinguishable on a CT scan and are likely to be missed or confused with image artifacts during radiological inspection. Therefore, even though the natural distribution of nodules in a general population contains far more small nodules than large nodules, not all of the small nodules are documented.

To illustrate this issue one may take a look at a histogram of nodule sizes in the subset of the Weill Cornell Medical Center database shown in Figure 7.1. Notice that the shape of the graph has a single distinct peak in a nodule size range between 2 and 3 mm. In theory, one might expect to see an increasing number of nodules to the left of the peak (as illustrated by the curve, connecting tops of the histogram bins), however few of them are actually documented. It is important to note that the position of the histogram peak would greatly depend on

image acquisition parameters, screening population and nodule documentation strategy; and that it will vary for different datasets and institutions.

For practical applications of detection systems, a minimum size threshold is usually specified where the visibility is reliable and where a specific action will be taken for a detected nodule of that size; that is, this minimum size depends on the clinical protocol that is supported by the detection system. Consequently, detection system developers report the performance with respect to a predefined size range [Sluimer et al., 2006; Enquobahrie et al., 2007; Zhang et al., 2005; Ge et al., 2005; Brown et al., 2002; Bae et al., 2005; Armato et al., 2005; Sahiner et al., 2007b]. In a clinical environment, the size of a nodule is usually estimated by an expert using one of the standardized measurement procedures; however, even for a single size metric, the disagreement between various readers may be quite large [Reeves et al., 2007a; Armato III et al., 2007; Reeves et al., 2007b].

The task for the detection system is to detect only nodules of a size above this threshold, and to disregard smaller objects. Thus, the output of the algorithm, which is usually represented by a set of detected nodule candidates, should be separated into two size categories according to the estimated size. However, the disagreement between automatically measured and ground truth size values alters the performance evaluation subset and, subsequently, changes reported detection performance. Further, it will be demonstrated how this disagreement affects both the detection sensitivity and the false positive rate of a detection

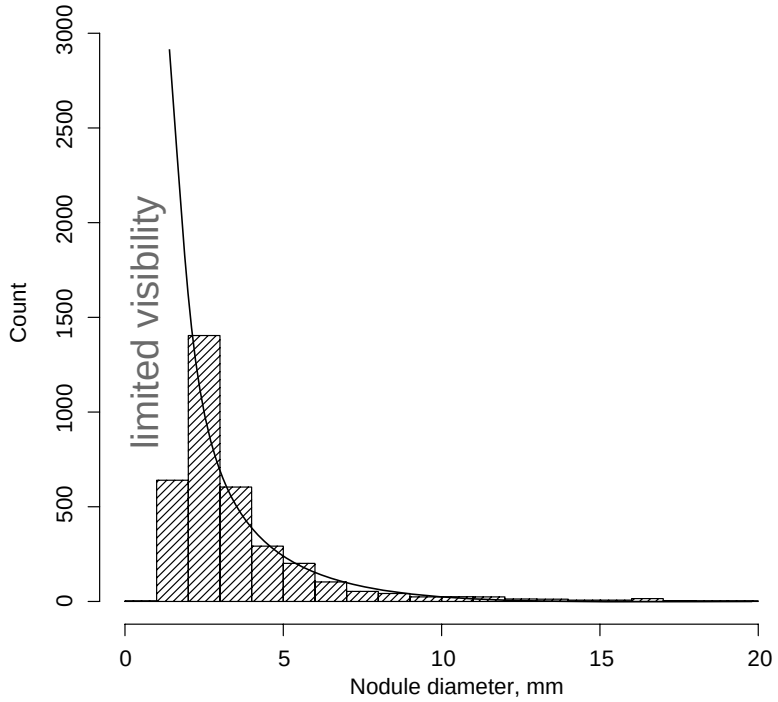


Figure 7.1: Histogram of diameters in Weill Cornell Medical Center database of 3503 nodules. The curve connecting the tops of histogram bars extrapolates to the small size range, showing the effect of limited visibility.

system.

7.1 Method of Δ-compensation

Given a specific size cut-off threshold, the difference between the nodule size estimation by a radiologist providing the ground truth and the automated candidate generation system complicates the procedure of detection performance evaluation. To measure the performance of the system for only nodules above certain threshold D_{th}, the detection algorithm needs to be configured to disregard smaller nodule candidates.

In the hypothetical situation illustrated in Figure 7.2, when there is no dis-

agreement between ground truth and automatic measurements, D_{th} clearly separates nodules and candidates into two size categories. Here the sensitivity $S_{x \geq D_{th}}$ and false positive rate $F_{x \geq D_{th}}$ of the system are easily calculated in terms of numbers of true and false candidates with the size x above the threshold D_{th}, while the smaller candidates are ignored:

$$S_{x \geq D_{th}} = \frac{N_{TP}}{N_{TP} + N_{FN}} = \frac{N_{TP}}{N_{x \geq D_{th}}}, \tag{7.1}$$

$$F_{x \geq D_{th}} = \frac{N_{FP}}{N_{cases}}, \tag{7.2}$$

where $N_{x \geq D_{th}}$ is the total number of "large" nodules and N_{cases} is the number of cases.

In a non-ideal world, the difference in size estimation leads to situations where some nodule sizes are either underestimated or overestimated. If a nodule of size $x_1 < D_{th}$ has corresponding candidate size $c(x_1) \geq D_{th}$, it must be considered as a false positive. Also, if a nodule of size $x_2 \geq D_{th}$ is measured as having size $c(x_2) < D_{th}$, it will be filtered out by the detection algorithm and recorded as a false negative. This effect is illustrated in Figure 7.3 and definitions of true and false positives and negatives are shown in Table 7.1. In this case, the sensitivity of the detection system can be calculated as (strokes denote variables related to the non-ideal situation):

$$S'_{x \geq D_{th}} = \frac{N'_{TP}}{N_{x \geq D_{th}}} = \frac{N_{TP} - N'_{FN(UE)}}{N_{x \geq D_{th}}}. \tag{7.3}$$

If we denote the probability for nodule of fixed size x to be underestimated in size as $P(c(x) < D_{th})$, we can express the number of false negatives due to size underestimate using the conditional mean:

$$N'_{FN(UE)} = N_{TP} \cdot E\left[P(c(x) < D_{th}) \,\middle|\, x \geq D_{th}\right]. \tag{7.4}$$

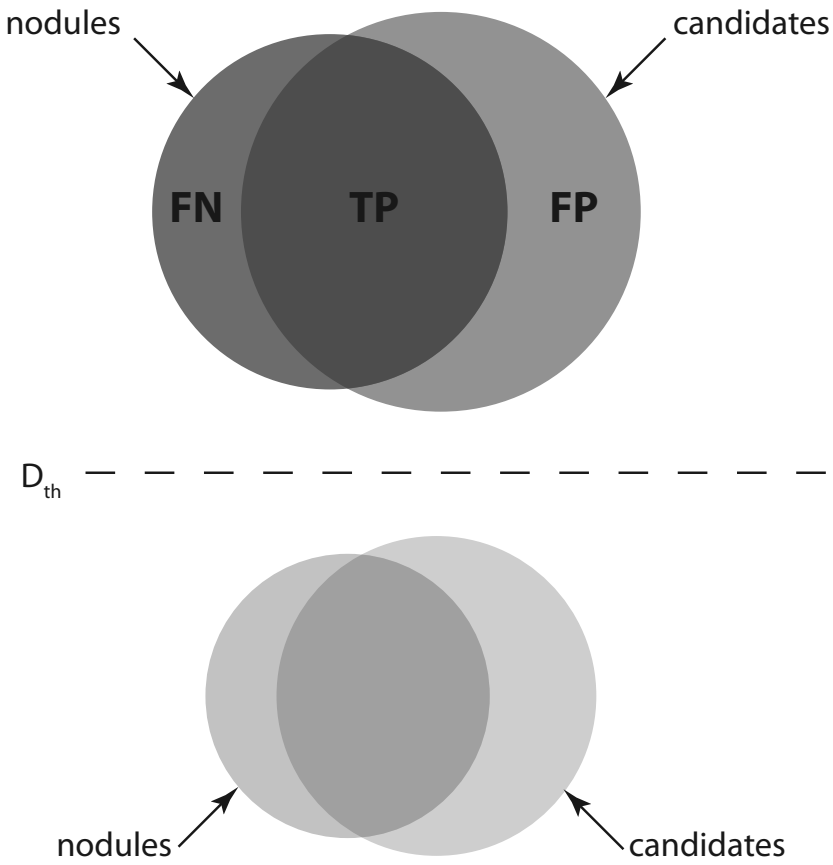


Figure 7.2: Ideal situation of zero size measurement error. Size threshold clearly separates nodules and candidates into two size categories.

After the substitutions, we get the following expression for the non-ideal sensitivity:

$$S'_{x \geq D_{th}} = S_{x \geq D_{th}} \cdot \left(1 - E\left[P(c(x) < D_{th}) \;\middle|\; x \geq D_{th}\right]\right). \tag{7.5}$$

Then, if we assume that for a nodule of size x there exists a random size measurement error $e(x)$ with the probability density function $f_{e(x)}$, so that $c(x) = x + e(x)$, we can find the probability of the size underestimate:

$$P(c(x) < D_{th}) = P(e(x) < D_{th} - x) = \int_{-\infty}^{D_{th}-x} f_{e(x)}(z)\, dz. \tag{7.6}$$

Finally, considering that the size of the nodule x is drawn from the distribu-

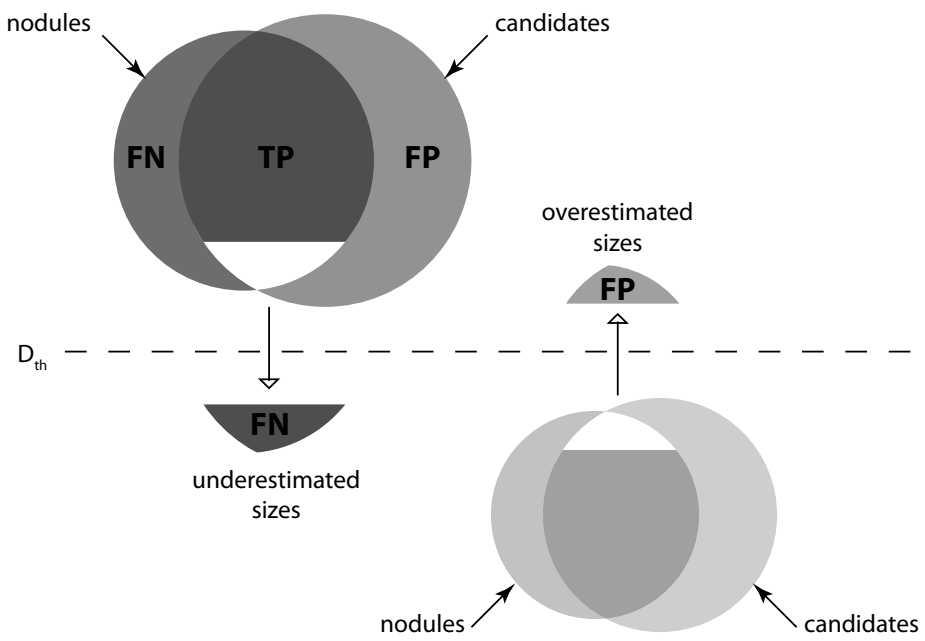


Figure 7.3: Diagram of nodule-candidate correspondence with the presence of non-zero size measurement error. Diameter threshold D_{th} separates nodules and candidates into two subsets; however because of the measurement error, some detected nodules have either underestimated or overestimated candidate size.

tion f_x, the value of the sensitivity in the presence of measurement error can be expressed as:

$$S'_{x \geq D_{th}} = S_{x \geq D_{th}} \cdot \left(1 - \frac{\int_{D_{th}}^{\infty} \int_{-\infty}^{D_{th}-x} f_{e(x)}(z)\, dz \cdot f_x(x)\, dx}{\int_{D_{th}}^{\infty} f_x(x)\, dx} \right). \tag{7.7}$$

Similarly, one can calculate the fraction of nodules overestimated in size that should be counted as false positives. Besides the natural distribution of nodules f_x, this fraction depends on the distribution of measurement error $f_{e(x)}$. An increase in this function variance decreases sensitivity and increases the false positive rate of the detection system.

There always will be discrepancy in size estimates due to real nodule size

Table 7.1: Definition of true and false positives for a nodule detection system with a minimum nodule size limit of D_{th}. "OE" and "UE" stand for size overestimate and underestimate, respectively.

		Nodule state		
		not exists	$x < D_{th}$	$x \geq D_{th}$
Candidate state	$c(x) \geq D_{th}$	FP	FP(OE)	TP
	$c(x) < D_{th}$	-	-	FN(UE)
	not exists	-	-	FN

uncertainty and critical differences in human and machine measurements. This discrepancy, if small, should not lessen reported performance of the detection system. For this reason, we propose a Δ-size tolerance range method to compensate for the size measurement error between a human expert and a computer system, in which we do not count the candidates with sizes in the immediate proximity of a cut-off threshold as either false positives or false negatives. This concept is illustrated in Table 7.2. Given a small specific value for Δ, a candidate is treated as overestimated or underestimated only if its size lies outside the interval bounded by points: $D_{th} - \Delta$ and $D_{th} + \Delta$. By applying this technique, we directly reduce the chances for a nodule and its corresponding candidate to be on opposite sides of the cut-off threshold, and increase reported sensitivity and reduce the false positive rate.

In the case of Δ-compensation, the probability of size underestimation with

Table 7.2: Compensation for size uncertainty using Δ radius for candidate sizes.

		Nodule state		
		not exists	$x < D_{th}$	$x \geq D_{th}$
Candidate state	$c(x) \geq D_{th} + \Delta$	FP	FP(OE)	TP
	$D_{th} \leq c(x) < D_{th} + \Delta$	FP	-	TP
	$D_{th} - \Delta \leq c(x) < D_{th}$	-	-	TP
	$c(x) < D_{th} - \Delta$	-	-	FN(UE)
	not exists	-	-	FN

compensation for uncertainty becomes:

$$P(c(x) < D_{th} - \Delta) = \int_{-\infty}^{D_{th}-x-\Delta} f_{e(x)}(z)\, dz, \tag{7.8}$$

and as Δ approaches infinity, the fraction of false negatives due to size underestimation reduces to zero. As a result, sensitivity of such a detection system is equal to one obtained with perfect ideal measurement:

$$S'_{x \geq D_{th},\ \Delta \to \infty} = S_{x \geq D_{th}}. \tag{7.9}$$

Generally, any value of Δ greater than the maximum difference in measurement error would result in the same (maximum) value of sensitivity. It can be shown that larger values of Δ have a favorable influence on the false positive rate of a detection system as well.

However, it is not desirable to use large values for Δ, since the candidate size information would be completely disregarded. Candidates that have large size measurement disagreement and clearly lie outside the size range of interest should be not be counted as true positives.

To illustrate the concept of Δ-compensation, let us consider a hypothetical situation of a detection system that detects nodules of the size 4 mm and larger. If the system generated a candidate of size 4.2 mm for a 3.8 mm nodule (as marked in the ground truth) this nodule is not counted as false positive, because human measurement could be done with error and, in fact, the nodule had the size of 4 mm. Similarly, when the system reports that a nodule has size 4.2 mm, while it was recorded in the ground truth as having size 3.8 mm, it is not counted as false negative. In the opposite situation, if a detection system reports a size of 4.2 mm on a 1 mm nodule, one must register a false positive, since the detection system has made a mistake and this nodule is clearly outside of the range of interest. Accordingly, if it reports 1 mm on a 4.2 mm nodule one must register a false negative. In order to implement this scheme, one may need to set a certain range of tolerance Δ for the difference between nodule and candidate sizes relative to the size cut-off threshold. In this case, the value $\Delta = 1\ mm$ would be appropriate.

7.2 Evaluation of the method

In order to evaluate the impact of size measurement disagreement on detection performance, an experimental automated nodule detection system that targets solid isolated nodules [Enquobahrie et al., 2007] was used. The evaluation set

consisted of 509 clinical cases from the CLDD 1.25 subset in which 690 solid isolated nodules were identified and measured by at least two expert radiologists. In the ground truth, nodule sizes were recorded as the average between the maximum axial diameter and its largest perpendicular. The size cut-off threshold was set to be 4 mm and the system was trained on the subset of 249 clinical cases containing 323 nodules. This system was then tested on the remaining 260 cases containing 367 nodules. A modified version of a performance evaluation procedure was used. It gave the flexibility to vary Δ and compute the resulting FROC over the test dataset. When $\Delta = 0$, the performance is the same as for the standard procedure that does not compensate for size measurement uncertainty. $\Delta = \infty$ results in the performance that would have been obtained by the detection system if the nodule size was measured perfectly without disagreement. The values of Δ in between corresponded to different degrees of compensation. Moreover, the operating point on the FROC curve corresponding to the maximum sensitivity of detection was selected. Then, the effect of Δ on the maximum sensitivity and false positive rate corresponding to this point was observed.

7.3 Impact of Δ-compensation

FROC curves were constructed for a documented dataset of 509 cases with the minimum size threshold set to 4 mm. Three curves, shown in Figure 7.4, correspond to different values of Δ compensation. The baseline performance corresponds to the case when there is no correction for size uncertainty, i.e. $\Delta = 0$. The best performance was recorded at $\Delta = \infty$, when candidate size was entirely ignored in performance evaluation. The middle curve corresponds to

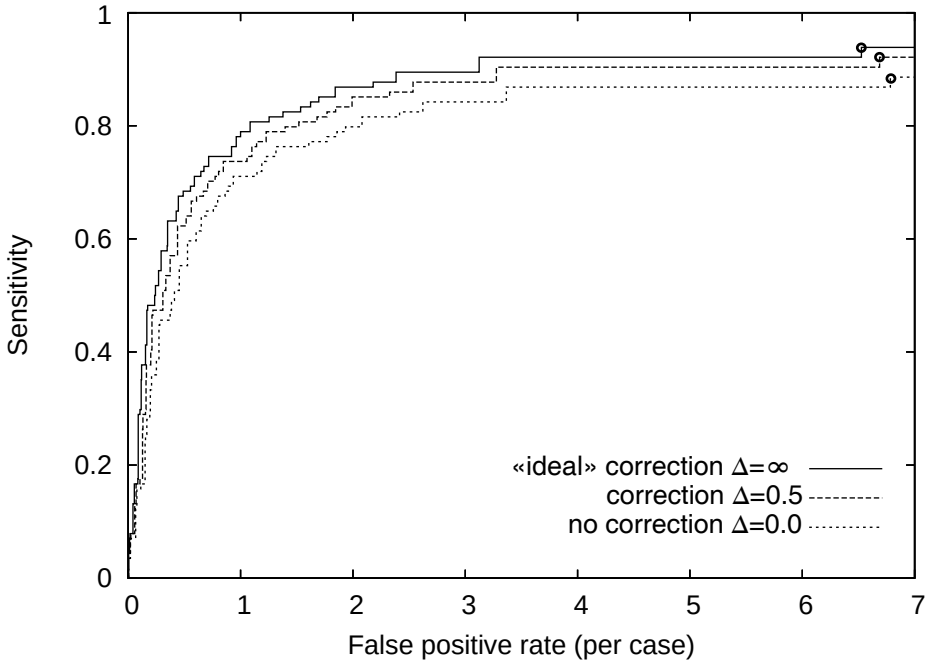
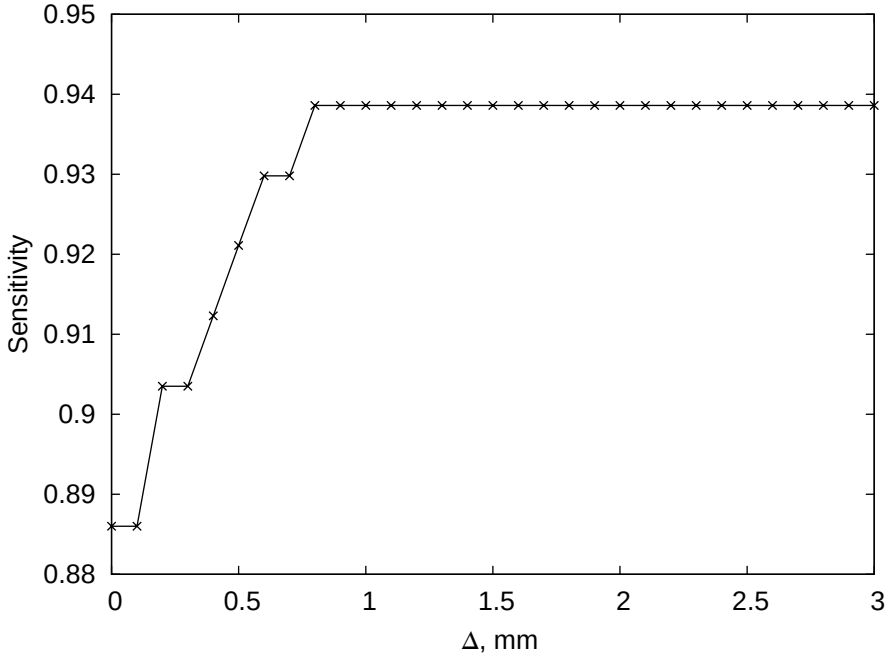


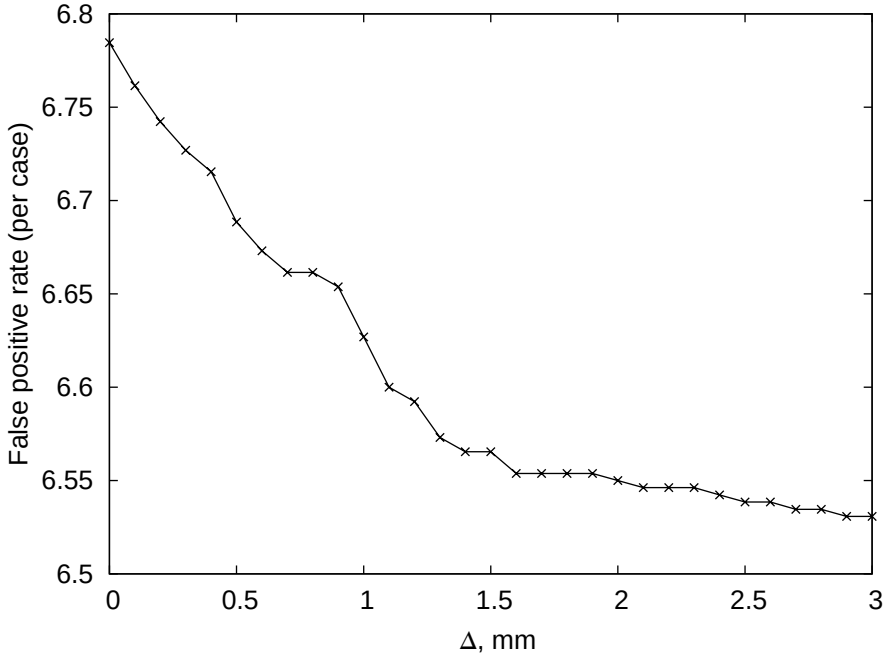
Figure 7.4: Effect of Δ on the FROC for detection of solid isolated nodules. The operating point, corresponding to the maximum sensitivity is denoted by circles.

$\Delta = 0.5\,mm$.

The influence of Δ on the sensitivity and specificity of the operating point, corresponding to maximum sensitivity is shown in Figure 7.5. The difference in nodule size estimate accounts for more than a 5% loss in sensitivity from 0.938 to 0.886 and a gain in false positive rate from 6.53 to 6.78. Any value of $\Delta > 0.8\,mm$ and $\Delta > 2.9\,mm$ did not help to improve sensitivity and false positive rate, respectively. These values directly relate the maximum extent of nodule size disagreement due to size underestimation and overestimation.



(a)



(b)

Figure 7.5: Effect of Δ on reported sensitivity (a) and false positive rate (b) of the detection system for the maximum sensitivity operating point.

7.4 Conclusion

Nodule size disagreement between human and automated measurements can have a significant impact on the performance evaluation of automated pulmonary nodule detection systems due to the minimum size cut-off employed by these systems. The study showed that this error reduced reported sensitivity by about 5% and increased the false positive rate by about 0.25 per case. We have presented a modified evaluation method that compensates for this error and generates detection performance close to the ideal case that would have been obtained if the nodule size was measured without error.

CHAPTER 8

TRAINING AND EVALUATION OF THE CAD SYSTEM

In previous chapters, the individual components of the CAD system were considered: the candidate generator, and the standard moments-based filter. The evaluation of these parts was done independently. The purpose of this section is to show how the system, described in Chapter 3, is trained and evaluated. One assumption made here is that the individual components of the system, namely lung segmentation, candidate generation, and feature computation algorithms, are already optimized and are given.

The organization of this chapter is as follows. First, a description of the available data partitioning for training and evaluation is provided. Then, we will proceed with the details of optimization of the detection cascade and selection of the classifier. An overview of the system performance on different datasets and nodule size ranges will conclude this chapter.

8.1 Partitioning of the datasets for training and evaluation

For the tasks of development and validation of a detection system, the traditional method of dividing available data into two independent training and test sets of cases was chosen. This way, the system can be optimized on a training subset and validated on a previously unseen test subset. Since in optimization of the image features, parameter decisions were already made using the data from the training subset, the techniques such as cross-validation were intentionally avoided to exclude any possible bias in subsequent performance evaluation. The partitioning of the dataset for training and evaluation is shown in

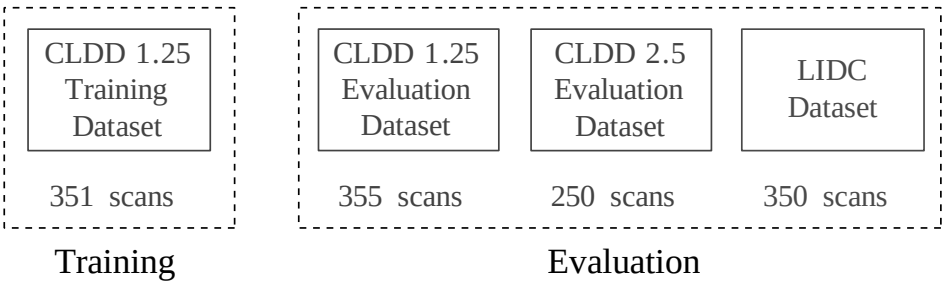


Figure 8.1: CT scan datasets used for training and performance evaluation of the CAD algorithm.

Figure 8.1. The enriched set of the CLDD 1.25 database was partitioned into two halves for training and testing using the parity of the numerical case identifier as the selection criterion. As the CLDD 2.5 and LIDC datasets were added later and have fewer cases, they were used only for system evaluation.

Candidates from the training cases with the corresponding set of computed features were used for optimization of serial and parallel false positive filters. Candidates from the evaluation sets were used to obtain the final FROC curves.

8.2 Optimization of serial FP reduction

Optimization of the serial FP reduction cascade was done as described in Chapter 3. The main purpose of the cascade was to sequentially reduce the number of candidates that are processed at a later stage of detection. With this architecture, processing of computationally expensive features is delayed until only a fraction of the original candidates remain, thus a significant speed-up of the algorithm was achieved.

During the training, each candidate was assigned either a $True$ or $False$ class

and the normalized distributions of feature values for both of the classes were analyzed. Based on these distributions, a threshold was selected such that all true positives are preserved and as many false positives as possible are filtered. Some of the true positives were considered $Outliers$ if the corresponding features had an extreme value (e.g. diaphragm nodules have an extremely high value for the total attachment feature, as discussed later in this chapter), and were ignored in determining the threshold.

Some of the features (contrast, distance to lung boundary), shown in Figure 8.2, did not provide clear separation of the classes and therefore were used only in the parallel classification stage.

The selected thresholds and the distribution of the values for the remaining features used in the cascade are shown in Figure 8.3. The candidates having a value greater than the selected threshold were disregarded and not considered further. An exception to this rule is the attachment features: the filtered candidates had feature values higher than the threshold.

Some parameters of the resulting candidate filtering cascade are shown in Table 8.1. Here the candidates greater than 4 mm were selected from the output of the candidate generator run on the test set. The average processing time per candidate and number of remaining candidates for each stage of the cascade are given. Note that attachment and proximity features were computed together at the same time to take advantage of reusing the same code.

The average processing time for the cascade organization is 88 seconds per case, which is $\tilde{4}.7$ times faster than the 412 seconds that are required for computing all the features without the cascade. This is achieved at the cost of reducing

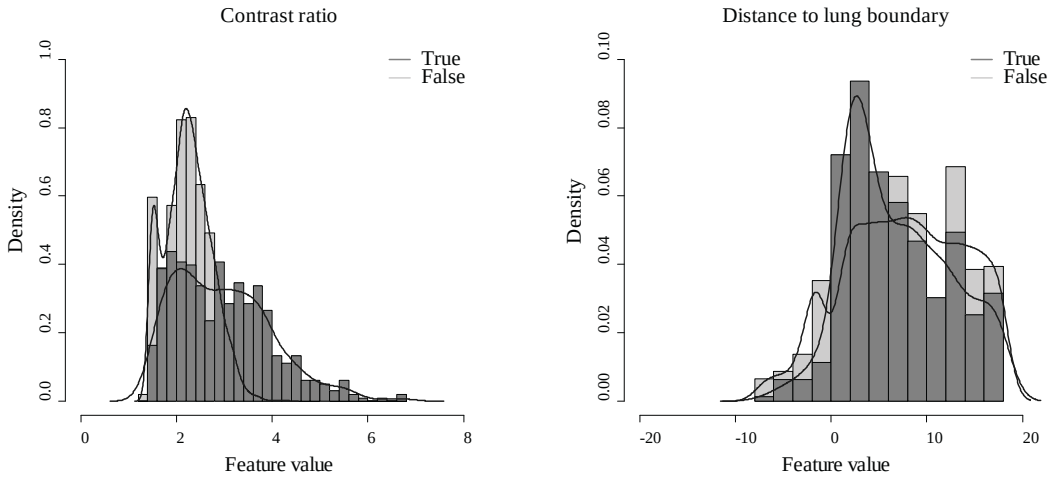


Figure 8.2: Distribution of contrast ratio and lung distance features: no single threshold will separate true candidates from the false candidates.

Table 8.1: Parameters of sequential false positive reduction cascade.

Feature	Average computation time (per candidate), s	Average candidates (per case)
Contrast	0.003	2285
Attachment	0.007	2285
Proximity	0.018	998
Shape	0.022	498
Moments-based	0.130	280

overall sensitivity by 6%.

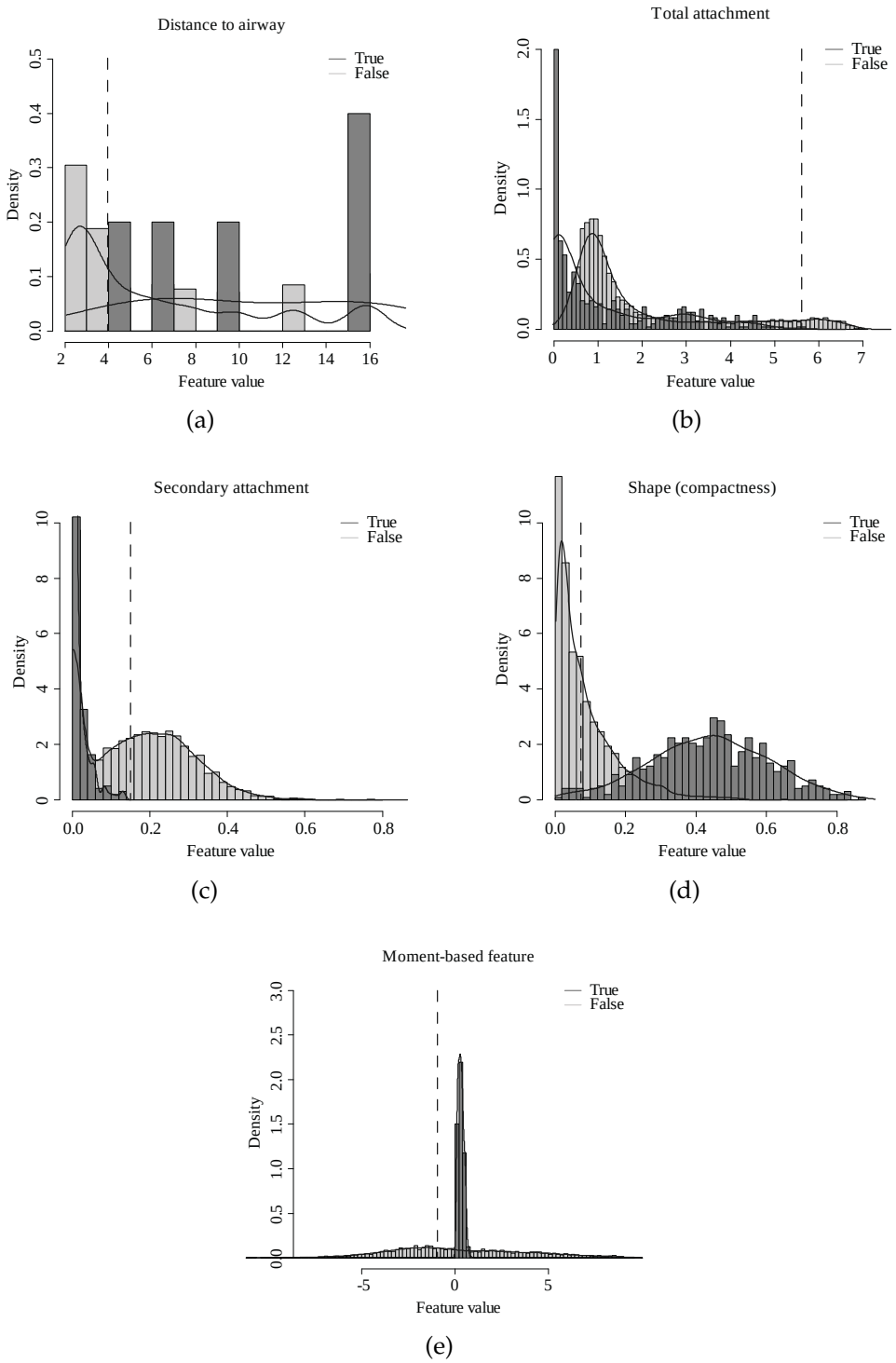


Figure 8.3: Distribution of features and selected thresholds (dashed line).

8.3 Training of the parallel classifier

As mentioned in Chapter 3, support vector machine (SVM) and distance weighted nearest neighbor (dwNN) classifiers were used in the stage of parallel candidate classification. These two methods use different approaches to construct the regression function, and therefore may provide different "opinion" and classification results. In training and evaluation, a subset of nodules with size greater than 5 mm from CLDD 1.25 dataset was considered.

The experiment showed only marginal advantage of the SVM over the dwNN classifier as shown in Figure 8.4. The quality of other system components such as candidate generators and false positive filters might have greater impact on the final performance of the system than does the selection of a particular classifier.

8.4 Evaluation of the system

In the set of experiments defined further in this section, certain properties of a detection system are explored.

As mentioned earlier, solid nodules occur much more frequently than nonsolid nodules; therefore, detection performance on solid nodules makes the largest contribution to the overall detection performance. For this reason and for the simplicity of the performance measurement, the majority of the following experiments were conducted on solid nodules only using the SVM as a final classifier. At the end of the section, however, combined performance of solid and nonsolid subsystems will be given.

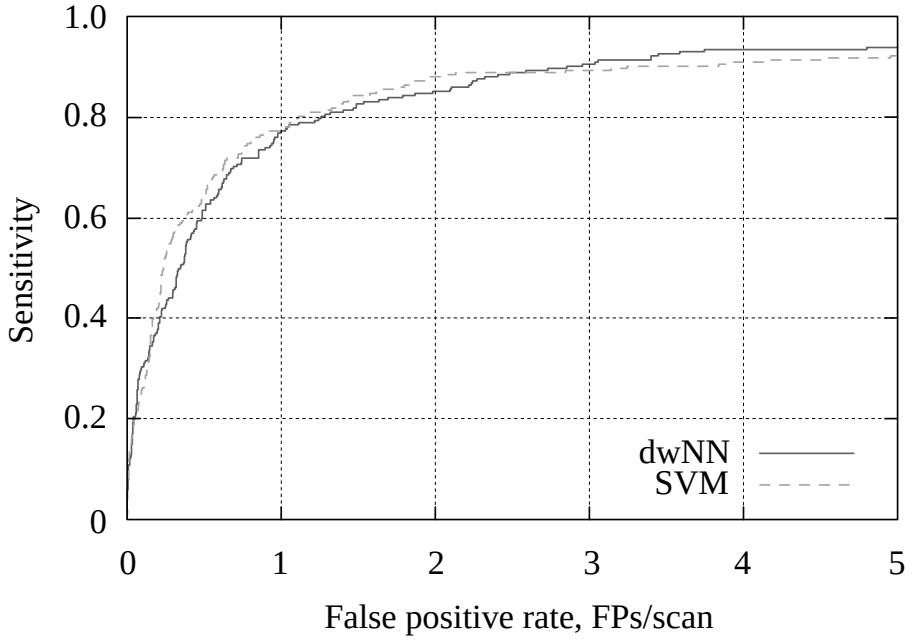


Figure 8.4: CLDD 1.25 set: detection performance on solid nodules of 5 mm in diameter and larger obtained using linear SVM and distance weighted nearest neighbor classifiers.

8.4.1 Effect of the target nodule size range

Depending on a clinical protocol, radiologists may be interested in identification of nodules of a certain size range. With increased nodule size, it is expected that coarse image resolution and image noise will have less impact on detection quality. Therefore, with fixed image resolution, an automated detection system may have better detection performance on nodules of larger sizes.

With this hypothesis in mind, the classifier was trained on the database using solid nodules by setting the minimum nodule size to be equal to 4 mm. The system was then tested on the subsets of nodules greater that 4, 5 and 6 mm in diameter, respectively. The CLDD 1.25 dataset was used.

The results of this experiment are shown in Figure 8.5. The target nodule size

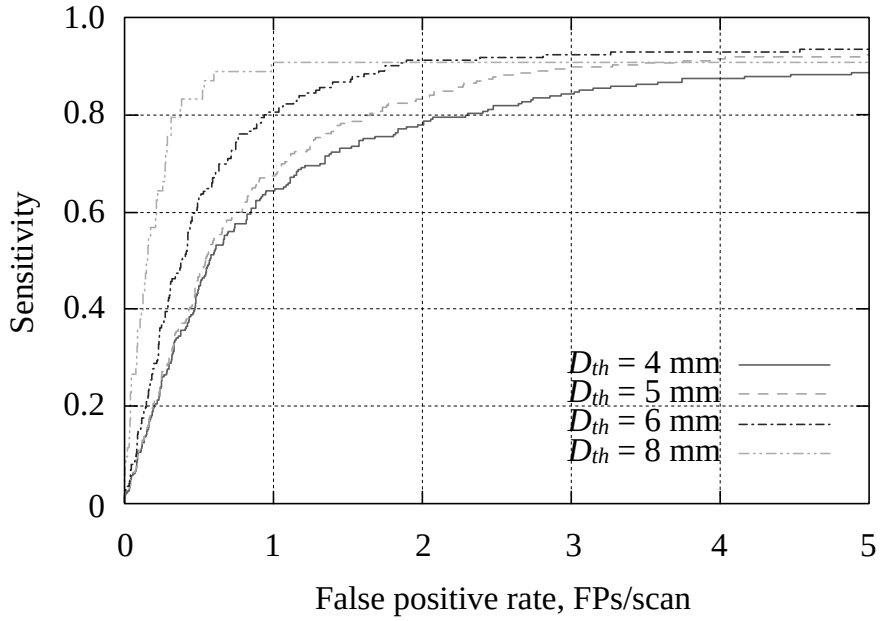


Figure 8.5: CLDD 1.25 set: detection performance of the solid detection subsystem trained on 4 mm nodules and tested on nodules of different size range.

range is an important factor affecting the performance. Smaller nodules have less variation in geometry and have much simpler attachment morphologies compared to larger nodules, and they are easier to detect. However, the number of candidates produced by the system increases roughly exponentially with the size as shown in Chapter 5, and setting the size cut-off threshold higher would reject many false positives. The experiment shows that while the sensitivity of the system is higher for smaller nodules, false positives contribute much more to the FROC curve and, as a result, performance on larger nodules is better.

8.4.2 Effect of the optimization for target nodule size range

In this experiment, we tested the hypothesis that training nodule size range would affect the performance on the test set, or in other words, whether it is reasonable to train the system for a particular size range. The size range of interest was set up to be 5 mm and the system was trained on nodules larger than 4 mm, larger than 5 mm, and larger than 6 mm, separately. Again the primary CLDD 1.25 dataset was used.

The experiment resulted in the curves shown in Figure 8.6. They indicate the positive impact of the training for the specific target nodule size range. For example, if one wants to disregard smaller nodules from consideration, the system can be trained on larger nodules and be expected to have better performance. This is explained by the fact that smaller nodules, especially on thick-slice images, are affected more by image noise and partial volume effect. Therefore, in striving for a better generalization, the classifier model settings may ”ignore” such properties of small nodules and perform more efficiently on larger nodules.

8.4.3 Effect of the slice thickness

There are several critical image parameters that may affect the performance of a detection system. Among these are image resolution and slice thickness. In the first experiment, to validate the system on the CT scans with different acquisition parameters, it was tested on the CLDD 1.25, CLDD 2.5 and LIDC datasets. Unlike the CLDD datasets, the LIDC dataset has scans of much higher variation in slice thickness. To minimize the effect of limited visibility, the minimum solid

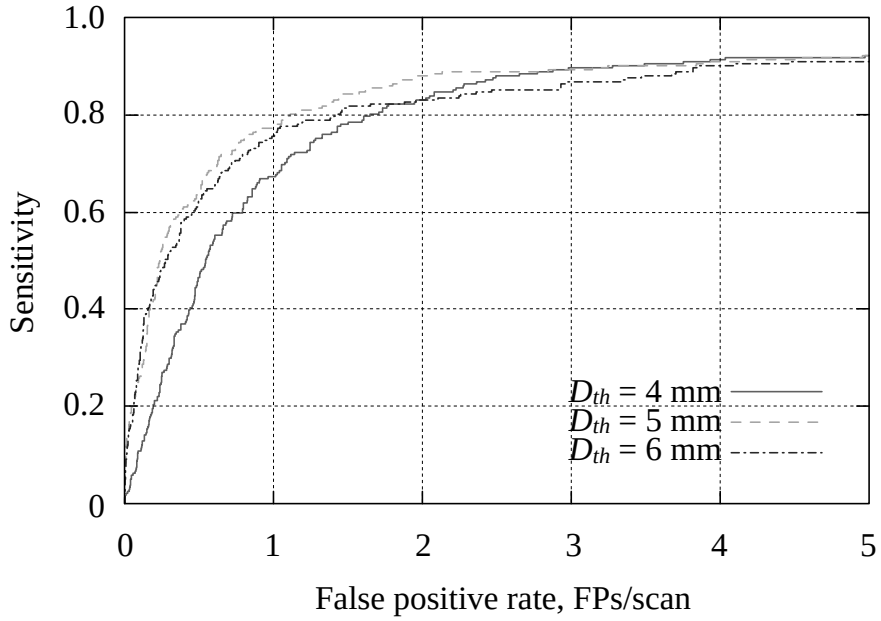


Figure 8.6: CLDD 1.25 set: detection performance on solid nodules of 5 mm and larger trained on subsets of different size range.

nodule size was selected to be 5 mm for all sets.

The FROC curves in Figure 8.7 indicate that the performance of the system is slightly degraded with the increase in slice thickness; however, from the comparison with Figures 8.5, 8.6, one may see that the performance of the system is affected more by target nodule size range. One must note that the difference in observed plots may also be explained by the population variation between the datasets.

8.4.4 Effect of windowing on detection of nonsolid nodules

In the next experiment the detection performance on nonsolid nodules with respect to sizes of 8 mm and greater was evaluated. Here the windowing thresh-

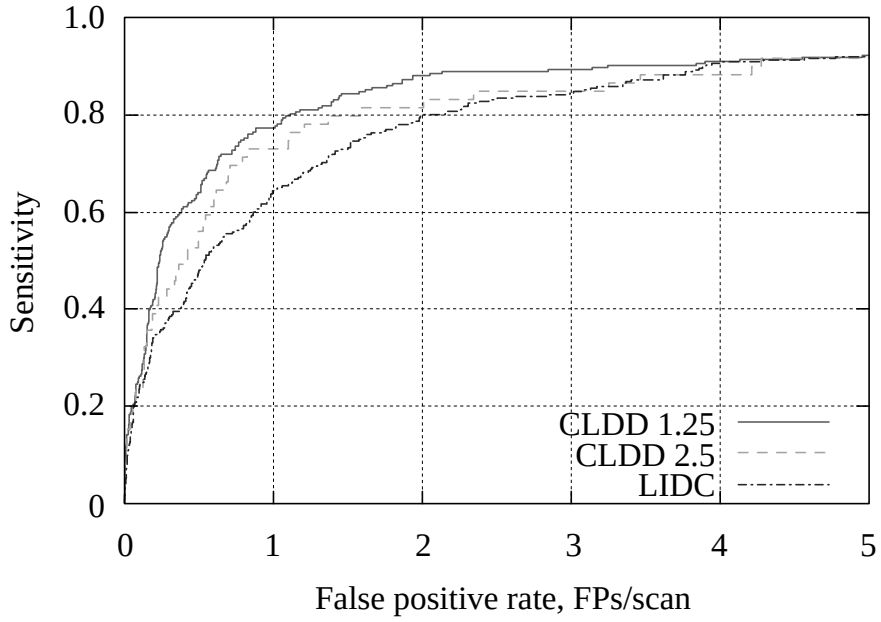


Figure 8.7: Detection performance of solid nodules obtained for the CLDD 1.25, CLDD 2.5 and LIDC test datasets with respect to the nodules of 5 mm in diameter and larger.

old T_{wind} of -680 HU as determined by [Browder, 2007] was used along with two additional levels of -630 HU, and -730 HU. The threshold resulting in better performance was selected for the evaluation of the system.

The three curves of the detection performance obtained on the training set corresponding to different threshold levels are plotted in Figure 8.8(a). The original threshold of -680 HU resulted in the highest possible sensitivity. The detection performance on the test set obtained for this threshold is illustrated by the FROC curve in Figure 8.8(b).

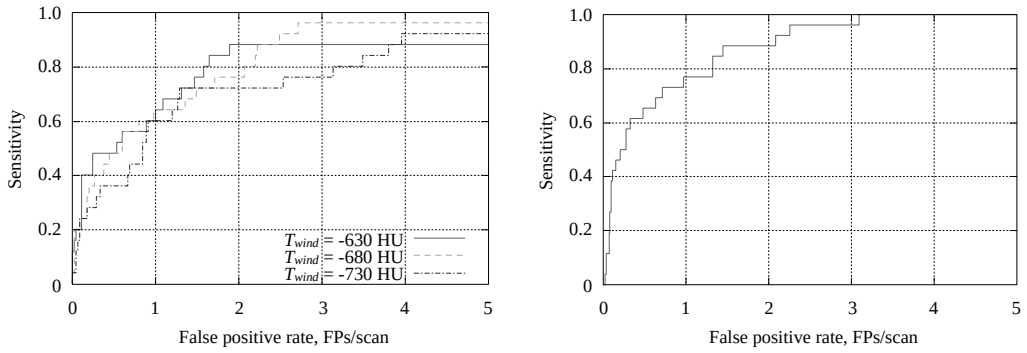


Figure 8.8: CLDD 1.25 set: detection performance on nonsolid nodules of 8 mm and greater. Shown are performances on the training set with respect to different windowing thresholds T_{wind} (a) and finalized performance on the test set with selected threshold $T_{wind} = -680HU$ (b).

8.5 Combined detection performance

Finally, the FROC curves from combining the results on the CLDD 1.25 dataset were constructed. Solid nodules equal to or larger than 5 mm and nonsolid nodules equal to or larger than 8 mm were considered. The results of the experiment are shown in Figure 8.9.

The shape of the combined detection performance graph is very similar to the one obtained for solid detection only. This was expected, and it can be explained by the relatively small number of nonsolid nodules that do not affect the overall sensitivity. A low false positive rate attained on solid nodules induces a shift to the right of the combined FROC plot.

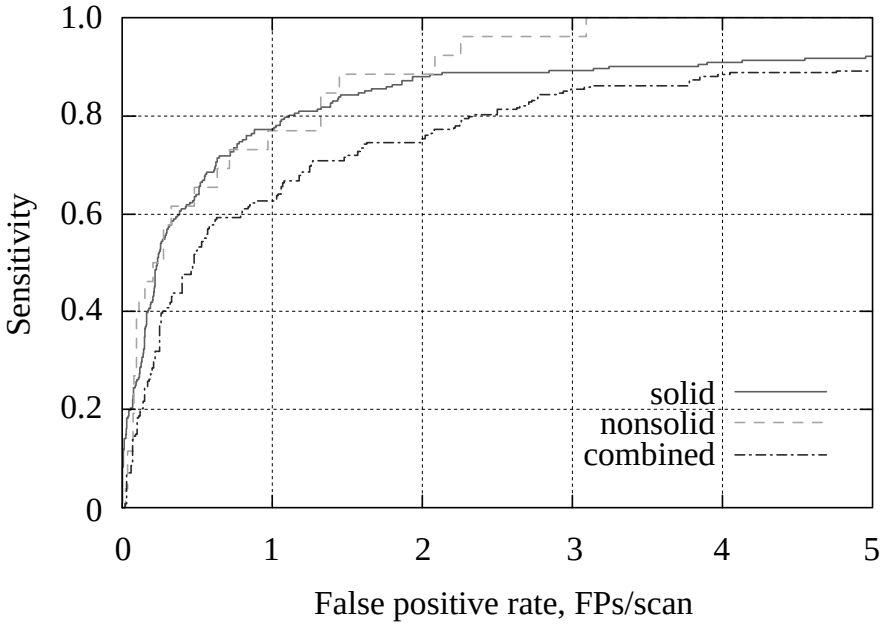


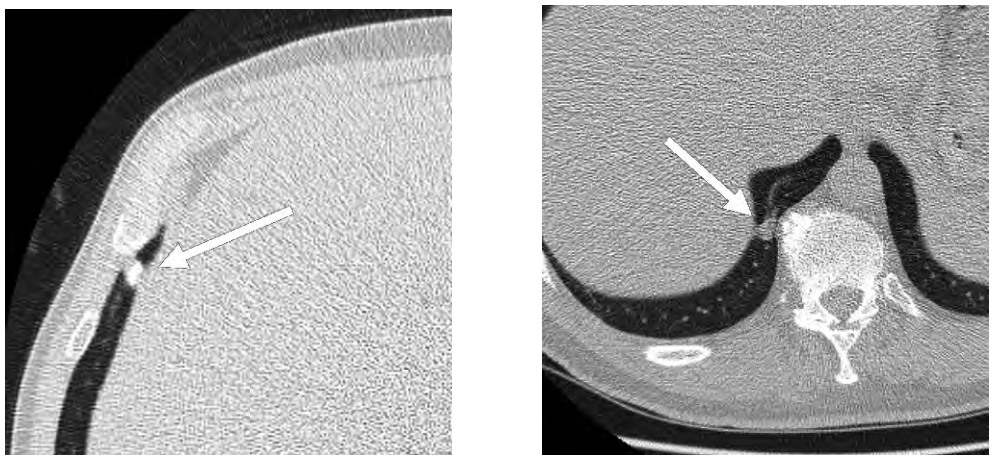
Figure 8.9: CLDD 1.25 set: combined detection performance on both solid and nonsolid nodules.

8.6 Discussion of the results

Independently of the operating point, a large fraction of the false negatives (missed solid nodules) for all datasets were in the thoracic diaphragm region of the lungs as shown in Figure 8.10. Although these nodules were identified by the candidate generator, none of the subsequent computed features were specifically designed to process nodules of these types. In general, any object within the lung with a large amount of attachment is difficult to identify. Two more examples of missed nodules are shown in Figure 8.11.

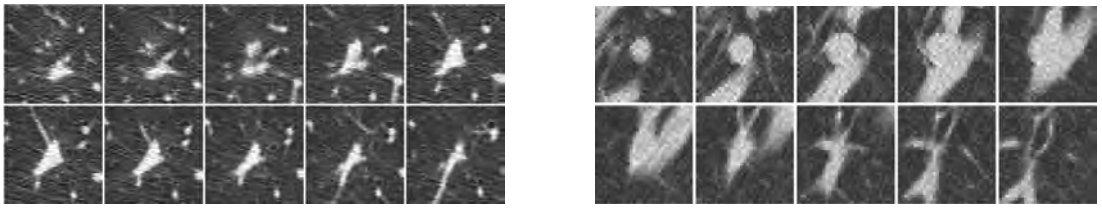
At high sensitivity levels, another source of false positives were the complex mediastinal vessel junctions and irregularities on the lung boundary as illustrated in Figure 8.12.

The main source of false positives findings in detecting nonsolid nodules



(a) (b)

Figure 8.10: Costophrenic (thoracic diaphragm region) nodules are among the most challenging false negatives.

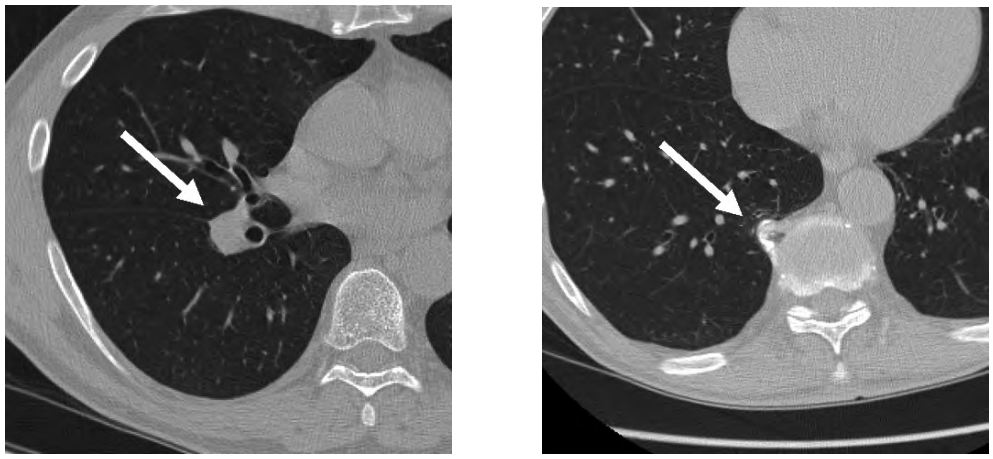


(a) (b)

Figure 8.11: False negatives: nodule confused with vascular branching point (a), complex attachment between a vessel and an airway(b).

were noisy parts of the lung parenchyma primarily near the thoracic diaphragm and heart, as illustrated by the example in Figure 8.13. Some CT scans were noisier than others, which led to a high variation in the number of such false positives across different scans.

Another source for increased false positive rate was the irregularities of larger lobulated nonsolid regions as displayed in Figure 8.14. Even though such a finding is not technically a false positive, an elegant scheme for counting these irregularities in the performance evaluation process was not developed.



(a) (b)

Figure 8.12: Common false positives: complex pulmonary vessel junction in mediastinal region (a) and osteophyte (b).

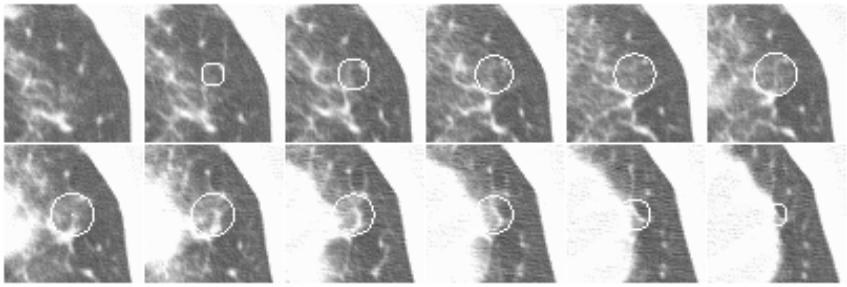


Figure 8.13: An example of a false positive in the thoracic diaphragm region of the lungs.

At low false positive rates, the nonsolid fraction of the candidates is high, which makes it possible to use this scheme together with a regular solid detection system. Both solid and nonsolid automated detection components of a system would provide a non-overlapping set of nodule candidates that can potentially lead to increased overall quantitative performance of the entire system.

This study has also demonstrated that in comparison to solid nodule detection schemes, identifying nonsolid nodules is easier: one does not need sophisticated techniques for reducing solid false positives that generally have much

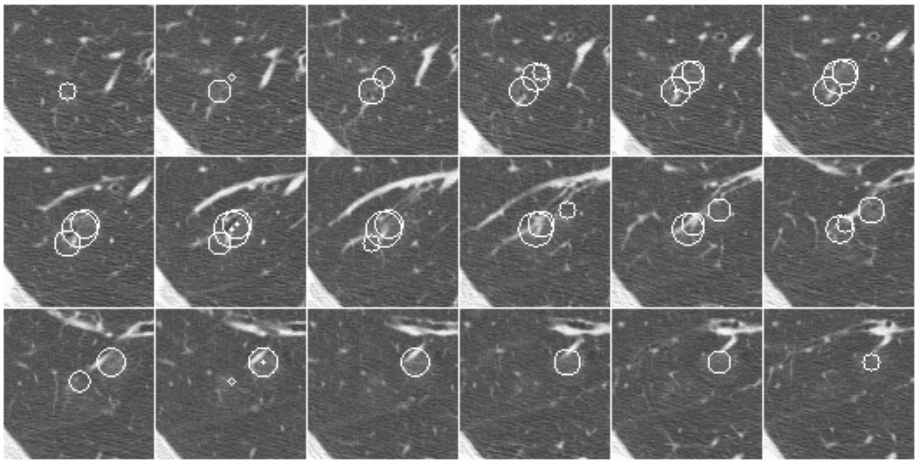


Figure 8.14: Example of a lobulated nonsolid lesion documented as a single nodule that resulted in four adjacent detected candidates.

higher variation in shape and location.

In spite of the absence of a preprocessing stage, the solid detection part requires more computational resources than nonsolid nodules. Therefore, when doing the detection in parallel, one needs to concentrate speed optimization efforts for only one of the subsystems.

Results from the detection of nonsolid nodules are better than the ones for solid nodules; there are very few objects within the lungs with a shape and consistency similar to nonsolid lesions. Moreover, the target nodule size range of interest is higher for nonsolid nodules, which helps to filter the majority of small false positives by size alone. In spite of the absence of extensive cascade and classifier optimization, our system can provide high sensitivity at false positive rates comparable to the results achieved by the best CAD systems reported in the literature.

The presented detection scheme achieved high detection performance at low false positive rates for both small solid and nonsolid nodule categories on dif-

ferent datasets and different size ranges of interest. The detection system may produce similar results independently of CT image acquisition parameters that confirm its robustness.

CHAPTER 9

CONCLUSION

Lung cancer, one of the deadliest diseases, needs to be diagnosed and treated in an early stage, when tumors are represented by small pulmonary nodules. Several clinical studies showed the effectiveness of using chest CT scans for nodule identification. As the nature of manual search is tedious and time consuming, the demand for computer-assisted methods is rising. The main focus of this dissertation was to advance the current practice of computer-aided detection of pulmonary nodules from CT scans.

One of the major accomplishments of this thesis is improved detection performance on both solid and nonsolid nodules compared to the results obtained by our research group in 2007 [Enquobahrie, 2007]. The progress in detection performance is shown in Figures 9.1 and 9.2. A detection sensitivity of approximately 90% was achieved at the rate of 2 FP per scan for solid nodules and 100% at the rate of 3.2 FP per scan for nonsolid nodules of clinically significant sizes. Such an improvement was the result of having the large, enriched CT scans dataset and the new advanced architecture of the system comprised of efficient computer algorithms.

The primary development dataset (CLDD 1.25) was increased in size to 706 CT scans compared to 250 scans in 2007. Thus the sizes of the development and evaluation subsets were increased proportionally with immediate implications for the increased generalization of detection algorithms. Moreover, the new dataset has a large population of solid and nonsolid nodules of clinically relevant sizes. This dataset can become a valuable resource for future development of the CAD system.

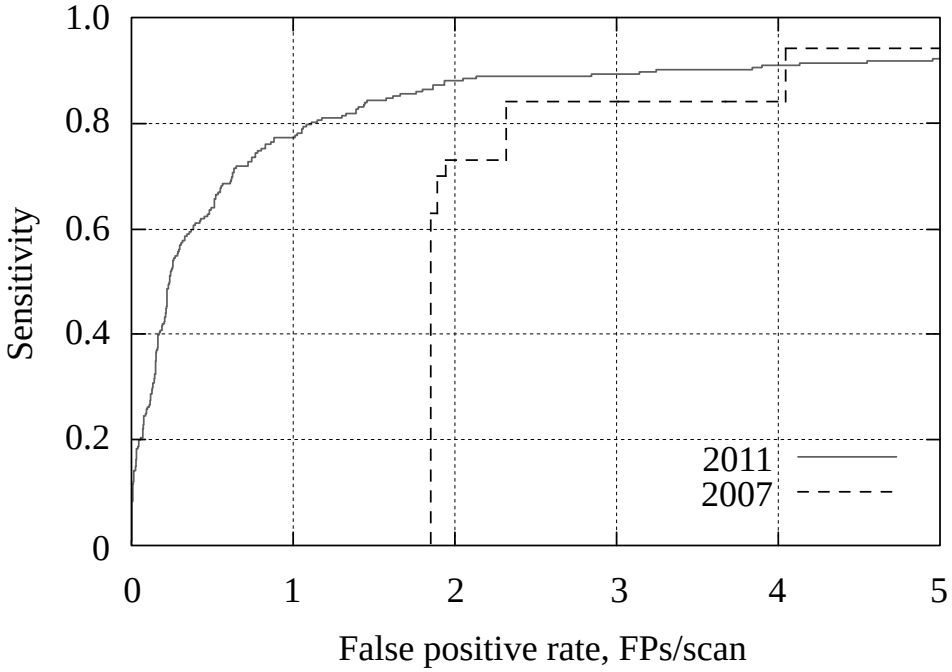


Figure 9.1: Progress of the CAD system performance with respect to the solid nodules of 5 mm in diameter and larger.

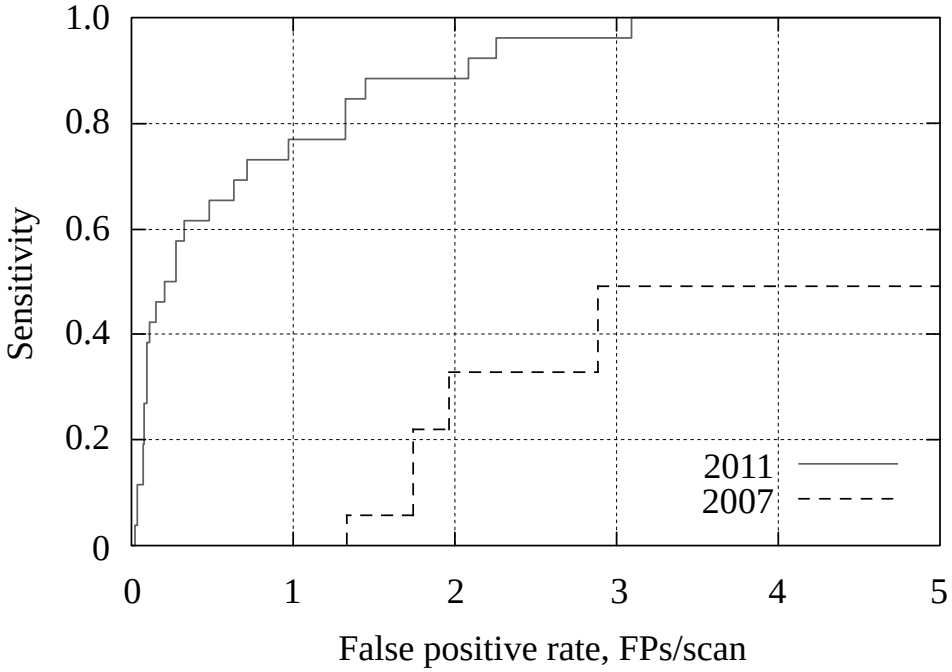


Figure 9.2: Progress of the CAD system performance with respect to the nonsolid nodules of 8 mm in diameter and larger.

Because of fundamental differences between solid and nonsolid nodules, the CAD system architecture was redesigned to incorporate parallel detection for both classes. Such organization helped to optimize the detection algorithms for each of the classes independently and to obtain better performance.

Both nodule detection modules were reinforced by a new candidate generator based on multiscale Laplacian of Gaussian filtering. The sensitivity of the candidate generator was shown to reach 0.998 (498/499) for solid nodules. A sensitivity of 1.000 (107/107) was achieved for nonsolid nodules after the application of the intensity windowing technique. Further, application of scale-normalized LoG filters allowed for both accurate size and nodule centroid estimation.

A standard moments based vessel bifurcation filter is another algorithmic improvement that targeted the most common type of false positive findings. The presented method resulted in 99% sensitivity and 80% specificity for the task of distinguishing nodules from vessel bifurcation points. Moreover, the filter was able to reject up to 69% of false positives with almost no loss in sensitivity after incorporating it into the design of the detection system.

In order to provide a more meaningful assessment of CAD system performance, we proposed a method of quantification of nodule size measurement error that affects the calculation of both the sensitivity and false positive rate. Our results suggest that the nodule size disagreement between human and automated measurements can have a significant impact on the detection performance evaluation. There is an associated error that reduced reported sensitivity by about 5% and increased the false positive rate by about 0.25 per case. We presented a modified evaluation method, called Δ-compensation, that accounts

for this error and generates detection performance close to the ideal case that would have been obtained if the nodule size was measured without error.

9.1 Future research

Future research in the field of automated nodule detection must be concentrated on approaching and exceeding the performance of a human reader. This will guarantee the improved quality of diagnosis, resulting in better prognosis for lung cancer patients.

Currently, automated detection systems, including the one reviewed in this thesis, can provide sensitivities above 90% (close to 100% after candidate generation step), which are higher than the sensitivity range of a human reader (30 – 70%). However, such levels are achieved at a high false positive rate, while expert radiologists operate at the false positive rates close to zero. Therefore, future CAD development must be primarily focused on reducing false positives to very low values while preserving the sensitivity as high as possible.

The candidate generator providing high sensitivity regardless of the nodule type is the first requirement for such a system. The second requirement is that the subsequent candidate classification should work in a way that allows preserving the sensitivity at approximately the same level while the false positive rate decreases. High variety of nodule presentations makes the design of the features that work universally for all nodule types extremely difficult. For example, diaphragm nodules and nodules adjacent to vessel branching points resulted in a very dissimilar and distinct set of features, which ultimately leaded to decreased system sensitivity at lower false positive rates.

For this reason, nodules of common shape and consistency should be aggregated into separate classes that are processed individually. Each class should have its own set of specifically tailored features. One obstacle to this scheme is that only a small number of samples may belong to a certain class, e.g. diaphragm nodules. From the algorithm learning perspective, it is very difficult to design a robust scheme that will generalize well if the training set is small. However, this can be addressed by enriching the dataset with these types and/or resorting to nodule simulation or insertion. Alternatively, particular rare nodule class can be ignored completely at the cost of reduced overall sensitivity.

As a result of the observations made during this research work, the nodule types that may require special attention are: diaphragm nodules, nodules with complex vessel attachments (both within parenchyma and attached to the wall), nodules of unusual shape (flat and elongated), and nonsolid nodules of very low contrast with lung parenchyma. The apperance of nodules in diseased lungs (emphysema, interstitial lung disease, chronic obstructive pulmonary disease) is also altered; and therefore, these nodules may deserve a special class as well.

As mentioned above, the major drawback of modern CAD systems is low sensitivity at the false positive rates close to zero. This may prompt switching from FROC to the use of alternative performance evaluation metrics such as sensitivity of detection at zero false positive rate. Systems optimized for this metric may have a greater chance of attaining and surpassing the radiologist at his native false positive rate. One advantage of such a metric is that combined output of multiple detectors cannot decrease it. Subsequently, the architecture of the corresponding detection system may incorporate a large number of the

simple detectors that are optimized for maximum specificity and that operate at a rate of zero false positives.

BIBLIOGRAPHY

G. Agam, S. G. Armato, and W. Changhua. Vessel tree reconstruction in thoracic CT scans with application to nodule detection. *IEEE Transactions on Medical Imaging*, 24(4):486– 499, 2005.

H. Arimura, S. Katsuragawa, K. Suzuki, F. Li, J. Shiraishi, S. Sone, and K. Doi. Computerized scheme for automated detection of lung nodules in low-dose computed tomography images for lung cancer screening. *Academic radiology*, 11(6):617–629, 2004.

S. G. Armato, A. S. Roy, H. Macmahon, F. Li, K. Doi, S. Sone, and M. B. Altman. Evaluation of automated lung nodule detection on low-dose computed tomography scans from a lung cancer screening program. *Academic Radiology*, 12(3):337–46, March 2005.

S. G. Armato III, G. McLennan, M. F. McNitt-Gray, C. R. Meyer, D. Yankelevitz, D. R. Aberle, C. I. Henschke, E. A. Hoffman, E. A. Kazerooni, H. MacMahon, A. P. Reeves, B. Y Croft, and L. P. Clarke. Lung Image Database Consortium: Developing a Resource for the Medical Imaging Research Community. *Radiology*, 232(3):739–748, 2004.

S. G. Armato III, M. F. McNitt-Gray, A. P. Reeves, C. R. Meyer, G. McLennan, D. R. Aberle, E. A. Kazerooni, H. MacMahon, E. J. R. van Beek, D. Yankelevitz, E. A. Hoffman, C. I. Henschke, R. Y. Roberts, M. S. Brown, R. M. Engelmann, R. C. Pais, C. W. Piker, D. Qing, M. Kocherginsky, and B. Y. Croft. The lung image database consortium (LIDC): An evaluation of radiologist variability in the identification of lung nodules on CT scans. *Academic Radiology*, 14(11): 1409–1421, 2007.

C. G. Atkeson, A. W. Moore, and S. Schaal. Locally weighted learning. *Artificial intelligence review*, 11(1):11–73, 1997. ISSN 0269-2821.

J. H. M. Austin, N. L. Muller, P. J. Friedman, D. M. Hansell, D. P. Naidich, M. Remy-Jardin, W. R. Webb, and E. A. Zerhouni. Glossary of terms for CT of the lungs: recommendations of the Nomenclature Committee of the Fleischner Society. *Radiology*, 200(2):327–332, 1996. ISSN 0033-8419.

K. Awai, K. Murao, A. Ozawa, M. Komi, H. Hayakawa, S. Hori, and Y. Nishimura. Pulmonary Nodules at Chest CT: Effect of Computer-aided Diagnosis on Radiologists' Detection Performance. *Radiology*, 230(2):347–352, 2004.

K. T. Bae, J. S. Kim, Y. H. Na, K. G. Kim, and J. H. Kim. Pulmonary nodules: automated detection on CT images with morphologic matching algorithm – preliminary results. *Radiology*, 236(1):286–93, 2005.

C. Bahlmann, X. Li, and K. Okada. Local pulmonary structure classification for computer-aided nodule detection. In *Medical Imaging 2006: Image Processing. Edited by Reinhardt, Joseph M. ; Pluim, Josien PW Proceedings of the SPIE*, volume 6144, pages 1775–1785, 2006.

H. A. Bastawrous, N. Nitta, and M. Tsudagawa. Reduction of False Positives in a CAD System for GGO Nodule Detection by Means of Neural Classification and CT Coronal View Examination. *IEEJ Transactions on Electronics, Information and Systems*, 127(6):922–927, 2007.

F. Beyer, L. Zierott, E. M. Fallenberg, K. U. Juergens, J. Stoeckel, W. Heindel, and D. Wormanns. Comparison of sensitivity and reading time for the use of

computer-aided detection (CAD) of pulmonary nodules at MDCT as concurrent or second reader. *European Radiology*, 17(11):2941–2947, 2007.

D. Blostein and N. Ahuja. Shape from texture: Integrating texture-element extraction and surface estimation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(12):1233–1251, 1989. ISSN 0162-8828.

K. W. Bowyer. Validation of medical image analysis techniques. *Handbook of Medical Imaging*, 2:567–607, 2000.

W. Browder. The segmentation of nonsolid pulmonary nodules in CT images. Master's thesis, Cornell University, 2007.

W. A. Browder, A. P. Reeves, T. V. Apananosovich, M. D. Cham, D. F. Yankelevitz, and C. I. Henschke. Automated volumetric segmentation method for growth consistency of nonsolid pulmonary nodules in high-resolution CT. In *Proc. of SPIE Vol*, volume 6514, pages 65140Y–1, 2007.

M. S. Brown, J. G. Goldin, R. D. Suh, M. F. McNitt-Gray, J. W. Sayre, and D. R. Aberle. Lung Micronodules: Automated Method for Detection at Thin-Section CT – Initial Experience. *Radiology*, 2002.

D. P. Chakraborty and L. H. Winter. Free-response methodology: alternate analysis and a new observer-performance experiment. *Radiology*, 174:873–81, 1990.

H. P. Chan, L. Hadjiiski, C. Zhou, and B. Sahiner. Computer-Aided Diagnosis of Lung Cancer and Pulmonary Embolism in Computed Tomography – A Review. *Academic Radiology*, 15(5):535–555, 2008.

P. Croisille, M. Souto, M. Cova, S. Wood, and et al. Pulmonary nodules: improved detection with vascular segmentation and extraction with spiral CT. Work in progress. *Radiology*, 197(2):397–401, 1995.

M. Das, G. Muhlenbruch, A. H. Mahnken, T. G. Flohr, L. Gundel, S. Stanzel, T. Kraus, R. W. Gunther, and J. E. Wildberger. Small pulmonary nodules: effect of two computer-aided detection systems on radiologist performance. *Radiology*, 241(2):564, 2006.

S. Diciotti, S. Lombardo, G. Coppini, L. Grassi, M. Falchini, and M. Mascalchi. The LoG Characteristic Scale: A Consistent Measurement of Lung Nodule Size in CT Imaging. *IEEE transactions on medical imaging*, 29(2):397, 2010.

M. Dolejsi and J. Kybic. Automatic Two-Step Detection of Pulmonary Nodules. In *Proceedings of SPIE*, volume 6514, page 65143J, 2007.

A. Enquobahrie, A. P. Reeves, D. F. Yankelevitz, and C. I. Henschke. Automated Detection of Small Solid Pulmonary Nodules in Whole Lung CT Scans from a Lung Cancer Screening Study. *Academic Radiology*, 14(5):579–593, 2007.

A. A. Enquobahrie. *Automated detection of pulmonary nodules from whole lung CT scans*. PhD thesis, Cornell University, 2007.

A. A. Farag, A. El-Baz, G. G. Gimel farb, R. Falk, and S. G. Hushek. Automatic detection and recognition of lung abnormalities in helical CT images using deformable templates. *Lecture Notes in Computer Science*, pages 856–864, 2004.

F. Fischbach, F. Knollmann, V. Griesshaber, T. Freund, E. Akkol, and R. Felix. Detection of pulmonary nodules by multislice computed tomography: improved detection rate with reduced slice thickness. *European radiology*, 13(10): 2378–2383, 2003.

H. B. Fleishon, M. Bhargavan, and C. Meghea. Radiologists' Reading Times Using PACS and Using Films: One Practice's Experience. *Academic radiology*, 13(4):453–460, 2006.

M. Frigo and S. G. Johnson. The design and implementation of FFTW3. *Proceedings of the IEEE*, 93(2):216–231, 2005. ISSN 0018-9219.

Z. Ge, B. Sahiner, H. P. Chan, L. M. Hadjiiski, P. N. Cascade, N. Bogot, E. A. Kazerooni, J. Wei, and C. Zhou. Computer-aided detection of lung nodules: False positive reduction using a 3D gradient field method and 3D ellipsoid fitting. *Medical Physics*, 32:2443, 2005.

C. I. Henschke, D. I. McCauley, D. F. Yankelevitz, D. P. Naidich, G. McGuinness, O. S. Miettinen, D. M. Libby, M. W. Pasmantier, J. Koizumi, N. K. Altorki, et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. *The Lancet*, 354(9173):99–105, 1999. ISSN 0140-6736.

C. I. Henschke, D. F. Yankelevitz, R. Mirtcheva, G. McGuinness, D. McCauley, and O.S. Miettinen. CT screening for lung cancer frequency and significance of part-solid and nonsolid nodules. *American Journal of Roentgenology*, 178(5): 1053–1057, 2002.

G. N. Hounsfield. Computerized transverse axial scanning (tomography): Part 1. Description of system. *British Journal of Radiology*, 46(552):1016, 1973.

T. Jiang and M. B. Merickel. Identification and boundary extraction of blobs in complex imagery. *Computerized Medical Imaging and Graphics*, 13(5):369–382, 1989. ISSN 0895-6111.

J. S. Jin and Y. Gao. Recursive implementation of LoG filtering. *Real-Time Imaging*, 3(1):59–65, 1997. ISSN 1077-2014.

A. C. Jirapatnakul, S. V. Fotin, A. P. Reeves, A. M. Biancardi, D. F. Yankelevitz, and C. I. Henschke. Automated nodule location and size estimation using a multi-scale laplacian of gaussian filtering approach. In *31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1028–1031, Sept. 2009.

T. Joachims. Making large-Scale SVM Learning Practical. Advances in Kernel Methods-Support Vector Learning, B. Scholkopf and C. Burges and A. Smola. 1999.

K. G. Kim, J. M. Goo, J. H. Kim, H. J. Lee, B. G. Min, K. T. Bae, and J. G. Im. Computer-aided Diagnosis of Localized Ground-Glass Opacity in the Lung at CT: Initial Experience. *Radiology*, 237(2):657–661, 2005.

J. J. Koenderink and A. J. van Doorn. Surface shape and curvature scales. *Image and Vision Computing*, 10(8):557–565, 1992.

Y. Lee, D. Y. Tsai, T. Hara, H. Fujita, S. Itoh, and T. Ishigaki. Improvement in automated detection of pulmonary nodules on helical x-ray CT images. In *SPIE*, volume 5370, pages 824–832, 2004.

Q. Li. Recent progress in computer-aided diagnosis of lung nodules on thin-section CT. *Computerized Medical Imaging and Graphics*, 31(4-5):248–257, 2007.

Q. Li, S. Sone, and K. Doi. Selective enhancement filters for nodules, vessels and airway walls in two- and three-dimensional CT scans. *Medical Physics*, 30(8): 2040–2051, 2003.

Q. Li, F. Li, and K. Doi. Computerized Detection of Lung Nodules in Thin-Section CT Images by Use of Selective Enhancement Filters and an Automated Rule-Based Classifier. *Academic radiology*, 15(2):165–175, 2008.

T. Lindeberg. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention. *International Journal of Computer Vision*, 11(4):283–318, 1993.

T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):79–116, 1998.

A. Lopes Pegna, G. Picozzi, M. Mascalchi, F. Maria Carozzi, L. Carrozzi, C. Comin, C. Spinelli, F. Falaschi, M. Grazzini, F. Innocenti, et al. Design, recruitment and baseline results of the ITALUNG trial for lung cancer screening with low-dose CT. *Lung Cancer*, 64(1):34–40, 2009. ISSN 0169-5002.

D. Marr. Early processing of visual information. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 275(942):483–519, 1976. ISSN 0080-4622.

D. Marr and E. Hildreth. Theory of edge detection. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, pages 187–217, 1980. ISSN 0080-4649.

K. Marten, C. Engelke, T. Seyfarth, A. Grillhosl, S. Obenauer, and E. J. Rummeny. Computer-aided detection of pulmonary nodules: influence of nodule characteristics on detection performance. *Clinical Radiology*, 60(2):196–206, 2005.

M. A. Meziane, R. H. Hruban, E. A. Zerhouni, P. S. Wheeler, N. F. Khouri, E. K. Fishman, G. M. Hutchins, and S. S. Siegelman. High resolution CT of the lung parenchyma with pathologic correlation. . *Radiographics*, 8(1):27, 1988. ISSN 0271-5333.

H. Miller. The FROC curve: a representation of the observer's performance for the method of free response. *The Journal of the Acoustical Society of America*, 46: 1473, 1969.

J. R. Muhm, L. R. Brown, and J. K. Crowe. Detection of pulmonary nodules by computed tomography. *American Journal of Roentgenology*, 128(2):267, 1977.

K. Murphy, B. van Ginneken, A. M. Schilham, B. J. de Hoop, H. A. Gietema, and M. Prokop. A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification. *Medical image analysis*, 13(5):757, 2009. ISSN 1361-8423.

J. Pu, B. Zheng, J. K. Leader, X. H. Wang, and D. Gur. An automated CT based lung nodule detection scheme using geometric analysis of signed distance field. *Medical Physics*, 35:3453, 2008.

A. P. Reeves and B. S. Wittner. Shape analysis of three dimensional objects using the method of moments. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 20–26, 1983.

A. P. Reeves, A. B. Chan, D. F. Yankelevitz, C. I. Henschke, B. Kressler, and W. J. Kostis. On measuring the change in size of pulmonary nodules. *Medical Imaging, IEEE Transactions on*, 25(4):435–450, 2006.

A. P. Reeves, A. M. Biancardi, T. V. Apanasovich, C. R. Meyer, H. MacMahon, E. J. R. van Beek, E. A. Kazerooni, D. Yankelevitz, M. F. McNitt-Gray, G. McLennan, S. G. Armato III, C. I. Henschke, D. R. Aberle, B. Y. Croft, and L. P. Clarke. The lung image database consortium (LIDC): A comparison of different size metrics for pulmonary nodule measurements. *Academic Radiology*, 14(12):1475–1485, 2007a.

A. P. Reeves, A. M. Biancardi, T. V. Apanasovich, C. R. Meyer, H. MacMahon, E. J. R. van Beek, E. A. Kazerooni, D. Yankelevitz, M. F. McNitt-Gray, G. McLennan, S. G. Armato III, D. R. Aberle, C. I. Henschke, E. A. Hoffman,

B. Y. Croft, and L. P. Clarke. The lung image database consortium (LIDC): Pulmonary nodule measurements, the variation and the difference between different size metrics. *SPIE International Symposium on Medical Imaging*, 2007b.

G. D. Rubin. Data explosion: the challenge of multidetector-row CT. *European journal of radiology*, 36(2):74–80, 2000. ISSN 0720-048X.

B. Sahiner, L. M. Hadjiiski, H. P. Chan, P. N. Cascade, E. A. Kazerooni, A. R. Chughtai, C. Poopat, et al. Effect of CAD on Radiologists Detection of Lung Nodules on Thoracic CT Scans: Analysis of an Observer Performance Study by Nodule Size. In *RSNA 93rd Scientific Assembly and Annual Meeting, Nov. 2007.*, page 265266, November 2007a.

B. Sahiner, L. M. Hadjiiski, H. P. Chan, J. Shi, T. Way, P. N. Cascade, E. A. Kazerooni, C. Zhou, and J. Wei. The effect of nodule segmentation on the accuracy of computerized lung nodule detection on CT scans: comparison on a data set annotated by multiple radiologists. *Proceedings of SPIE*, 6514:65140L, 2007b.

E. G. Schaner, A. E. Chang, J. L. Doppman, D. M. Conkle, M. W. Flye, and S. A. Rosenberg. Comparison of computed and conventional whole lung tomography in detecting pulmonary nodules: a prospective radiologic-pathologic study. *American Journal of Roentgenology*, 131(1):51, 1978.

A. M. R. Schilham, B. van Ginneken, and M. Loog. Multi-scale Nodule Detection in Chest Radiographs. *LECTURE NOTES IN COMPUTER SCIENCE*, pages 602–609, 2003.

I. Sluimer, A. Schilham, M. Prokop, and B. van Ginneken. Computer analysis of computed tomography scans of the lung: a survey. *IEEE Transactions on Medical Imaging*, 25(4):385–405, April 2006.

B. Van Ginneken. Computer-Aided Diagnosis in Thoracic Computed Tomography. *Imaging Decisions MRI*, 12(3):11–22, 2008.

V. N. Vapnik. *The nature of statistical learning theory*. Springer Verlag, 2000. ISBN 0387987800.

P. Viola and M. Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. In *Proceedings of CVPR2001*, volume 1, 2001.

D. Wormanns, F. Beyer, S. Diederich, K. Ludwig, and W. Heindel. Diagnostic performance of a commercially available computer-aided diagnosis system for automatic detection of pulmonary nodules: comparison with single and double reading. . *RoFo: Fortschritte auf dem Gebiete der Rontgenstrahlen und der Nuklearmedizin*, 176(7):953, 2004.

D. M. Xu, H. Gietema, H. de Koning, R. Vernhout, K. Nackaerts, M. Prokop, C. Weenink, J. W. Lammers, H. Groen, M. Oudkerk, et al. Nodule management protocol of the NELSON randomised lung cancer screening trial. *Lung Cancer*, 54(2):177–184, 2006.

M. Yang, S. Periaswamy, and Y. Wu. False Positive Reduction in Lung GGO Nodule Detection with 3D Volume Shape Descriptor. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 1, 2007.

X. Ye, X. Lin, G. Beddoe, and J. Dehmeshki. Efficient Computer-Aided Detection of Ground-Glass Opacity Nodules in Thoracic CT Images. In *Engineering in Medicine and Biology Society, 29th Annual International Conference of the IEEE*, pages 4449–4452, 2007.

R. Yuan, P. M. Vos, and P. L. Cooperberg. Computer-aided detection in screening

CT for pulmonary nodules. *American Journal of Roentgenology*, 186(5):1280–1287, 2006.

X. Zhang, G. McLennan, E. A. Hoffman, and M. Sonka. Automated detection of small-size pulmonary nodules based on helical CT images. *Proceedings of Information Processing in Medical Imaging: 19th International Conference, IPMI*, 2005.

B. Zhao. Automatic detection of small lung nodules on CT utilizing a local density maximum algorithm. *Journal of applied clinical medical physics*, 4(3): 248, 2003. ISSN 1526-9914.

F. Zhao, P. R. S. Mendonca, R. Bhotika, and J. V. Miller. Model-based junction detection with applications to lung nodule detection. *ISBI. (April 2007)*.