16781700

general fregerikken og en en det skreverske som after after at hande til

A CONTEXT BASED AUTOMATED SYSTEM FOR LUNG NODULE DETECTION IN CT IMAGES

by

Maciej Dajnowiec B.A.Sc., University of Toronto, 2003

A Thesis presented to Ryerson University in partial fulfillment of the requirements for the degree of

Master of Applied Science

in the Program of Electrical and Computer Engineering

Toronto, Ontario, Canada, 2005

© (Maciej Dajnowiec) 2005

PROPERTY OF RYERSON UNIVERSITY LIDRARY

UMI Number: EC53014

All rights reserved

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI®

UMI Microform EC53014 Copyright 2008 by ProQuest LLC All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

> ProQuest LLC 789 East Eisenhower Parkway P.O. Box 1346 Ann Arbor, MI 48106-1346

I hereby declare that I am the sole authour of this thesis.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Signature

I further authorize Ryerson University to lend this thesis by photocopying or other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

the state of the s

ense anestro de las quille

autor

ii

Signature

Ryerson University requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

ngangan shi na 12

a particular a la companya de la com

a na nagazina politika na katalaran gazing da Areada ya shekara na katala

지 않는 것은 것 같아.

iii

A CONTEXT BASED AUTOMATED SYSTEM FOR LUNG NODULE DETECTION

IN CT IMAGES

Master of Applied Science

2005

Maciej Dajnowiec

Electrical and Computer Engineering

Ryerson University

Abstract

This thesis is focused on automatic lung nodule detection in CT images. CAD systems are suited for this task because the sheer volume of information present in CT data sets is overwhelming for radiologists to process. The system developed in this thesis presents a fully automatic solution that applies a sequential algorithm which strongly focuses on nodule context. The system operates at a rate of 80% sensitivity with 3.05 FPs per slice. Our testing data, consisting of 19 CT data sets containing 239 lung nodules, is extremely robust when compared with other documented systems. In addition it introduces many new approaches such as a tight bounding, vessel connectivity, perimeter analysis, adaptive MLT and region growing based lung segmentation. The experimental results produced by this system are an affirmation of the competitiveness of its performance when compared to other documented approaches.

Acknowledgments

I would like to first thank Professor Javad Alirezaie for taking a chance on me and accepting me into the program. I also want to thank him for his guidance and all the work he has put into helping me produce this thesis, it would not have been possible without him.

I would like to thank Dr. Paul Babyn for providing us with the CT data we needed for our experiments. I want to thank him for all the time he spent identifying lung nodules and giving us insight into them without which it would have been immeasurably more difficult to develop the system that we did.

I would like to thank my lab mates throughout my tenure here; Mahdi, Nezam, Adeel, Peyman, Rushin, Negar, Jacob and Masha. I could not have asked for nicer people and a warmer environment within which to get all of this done.

I would like to thank my examining committee members; Dr. Anpalagan, Dr. Androutsos and Dr. Zhang for being a part of the home stretch in this process.

Lastly, I would like to thank Professor Phil Anderson and Professor Tarek Abdelrahman who were kind enough to write letters for me that must have been kind since I was granted the opportunity to do what I have completed now.

v

Dedication

To my guardian angel Candice and the children we have yet to raise, my father, Waclaw, my mother Barbara and my big brother Mateusz. Thanks to all of you I will always be the most fortunate man in the world.

化合金合金 化合物管理合金合金管理合金

และ รูกษณฐาวกลังสร้างได้ และ วัดการ แล้วสังได้ได้ได้ได้ได้ได้ได้ การรู้ได้ และ วายสร้างการ การการการการ การการกับการสะดาชได้ หรุงเพราะที่แก่ การให้นี้ได้ การการกับการสะดาชได้ หรุงเพราะที่แก่ การให้นี้

an a tha a tha an tha an the state the state and a

Table of Contents

entreport in the state of the property and

	(1,1,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2	Σ.
1 Introduction	and the states of project	1
1.1 CT Imaging		1
1.2 Cancer Statistics		4
1.3 Lung Nodules		6
1.4 CAD System		10
1.5 Outline		11
2 Literature Survey		12
2.1 LDA Based Techniques		12
2.2 Rule Based Techniques		14
2.3 Template Matching Techniques		16
2.4 Fuzzy Neural Network Technique		17
2.5 Hybrid Rule Based and LDA systems		19
2.6 Automatic Lung Segmentation System		20
2.7 Comparing Results		21
3 Lung Nodule Synthesis		24
3.1 Lung Nodule Synthesis Approach		24
3.2 Lung Nodule Synthesis Results		29
4 Lung Volume Segmentation		32
4.1 Lung Volume Thresholding		33
4.2 3D Lung Volume Extraction		39
5 Lung Nodule Detection		52
5.1 Lung ROI Extraction		52
5.2 FP reduction		61

6 Experimental Results	79
6.1 Lung Segmentation	79
6.2 ROI Extraction Results	82
6.3 Lung Nodule Detection Results	85
7 Conclusions	
7.1 Future Work	103

Bibliogrpahy

¢

106

11. 1 2 1

ju i er veret

onge det hande ande ander det

en gerinden gerinden die Staar van die St

and read Neural Metrock Cooking

1.5 Hybrid Ente Based and LDA symone

Constant a charactering and a state and the second of a

Alter Briangerold Ca

anteneargié analest contra

genter open to and a spect

annach na dhairtean a t

en al statul a sel sel a part a con

List of	f Tables		
Table 1.1	2005 US statistics on expected new cancer cases	4	
Table 1.2	2005 US statistics on expected cancer deaths	5	*
Table 1.3	2005 US statistics on 5 year survival rates for different cancers	5	4
Table 2.1	Nodule detection system performance in subsequent testing	22	
Table 5.1	Grey level values of different CDF values	54	
Table 5.2	Grey level values at 65%, 92.5% and 97.5% CDF values	59	
Table 5.3	Feature thresholds for different types of ROIs	72	
Table 6.1	Comparison between lung segmentation approaches	80	
Table 6.2	Simulated nodule testing results	85	;
Table 6.3	Phase 1 lung nodule detection results	87	, J
Table 6.4	Phase 2 lung nodule detection results	88	
Table 6.5	Phase 3 lung nodule detection results	89	
Table 6.6	Phase 4 lung nodule detection results	90	
Table 6.7	Performance comparison of lung nodule detection techniques	92	
Table 6.8	Effect of CT slice resolution on appearance of small nodules	97	

!

ix

1.2

List of Figures

. .

						: .	
Figure 1.1	Thoracic CT slice	•.				2	
Figure 1.2	Frontal radiogram	2		ing provide	• •	3	
Figure 1.3	Juxtapleural nodules	£3 *				² 7	. '
Figure 1.4	Ground glass nodules					7	
Figure 1.5	Noisy nodule					8	
Figure 1.6	Diaphragm adjacent nodules					8	
Figure 1.7	Large lung wall nodule					9	
Figure 3.1	A CT scan with a nodule in the righ	it lung				26	
Figure 3.2	Enlarged image of lung nodule in ri	ght lung	1			26	
Figure 3.3	3D mesh of a lung nodule					27	
Figure 3.4	Nodule size distribution within a C	Γ data se	et			28	
Figure 3.5	A real and synthetic nodule in the sa	ame CT	slice			29	
Figure 3.6	3D mesh of a synthetic nodule					30	
Figure 3.7	Screenshot of lung nodule synthesis	s GUI				31	
Figure 4.1	Nodule detection subsystem diagram	n				32	
Figure 4.2	A CT slice with the valid portion id	entified				34	
Figure 4.3	The HU distribution of a high inten	sity CT :	scan			34	
Figure 4.4	Thresholding comparison of a high	intensity	y CT slice	e		36	
Figure 4.5	A CT slice with normal contrast					37	
Figure 4.6	The HU distribution of a normal CT	Г scan				37	
Figure 4.7	Thresholding comparison of a regul	lar intens	sity CT sl	ice		38	
Figure 4.8	Eight pixel connectivity					39	
Figure 4.9	Thresholding a regular intensity CT	slice				40	
Figure 4.10) Removing the background non-boo	dy pixels	s from a C	CT slice		40	
Figure 4.11	l Removing body pixels from the lui	ngs				41	
Figure 4.12	2 Lung volume splitting due to diaph	ragm in	trusion			43	
Figure 4.13	3 A CT slice with the lung separated	by vasc	ular tissu	e		43	

Figure 4.14 Threshold of a lung volume split by vascular tissue	44
Figure 4.15 A CT slice near the top of the lungs	44
Figure 4.16 Thresholding and volumetric elimination near the top of the lungs	45
Figure 4.17 A CT slice with many objects attached to the lung walls	45
Figure 4.18 A threshold of a CT slice with a jagged border	46
Figure 4.19 Effect of morphological closing on the lung walls	47
Figure 4.20 An examination of morphological closing on the lung walls	48
Figure 4.21 A comparison of a pre/post morphology segmented CT slice	48
Figure 4.22 A summary of our lung segmentation process	51
Figure 5.1 CDFs of three different lung volumes	53
Figure 5.2 CT slices from three different lung volumes	54
Figure 5.3 A single CT slice thresholded at multiple CDF values	55
Figure 5.4 A single CT slice thresholded at multiple CDF values	56
Figure 5.5 A single CT slice thresholded at multiple CDF values	57
Figure 5.6 A CT slice thresholded at 65%, 92.5% and 97.5% CDF values	59
Figure 5.7 A CT slice thresholded at its 62.5% CDF value	60
Figure 5.8 A disc structuring element of radius 1	63
Figure 5.9 Intensity threshold eroded at different thresholds	63
Figure 5.10 A disc structuring element of radius 2	64
Figure 5.11 Normally oriented ellipses with traditional bounding boxes	65
Figure 5.12 A comparison between a tight and normal bounding box	66
Figure 5.13 A CT slice containing a border case nodule	68
Figure 5.14 A threshold of a CT slice containing a border case nodule	69
Figure 5.15 A close up of a thresholded border case nodule	69
Figure 5.16 A grey level contrast spectrum	73
Figure 5.17 The concept of nodule trimming and compaction	74
Figure 5.18 A CT slice containing different types of vertical vessels	75
Figure 5.19 Vertical and horizontal vessel connectivity	76
Figure 5.20 Summary of our FP reduction process	78

Figure 6.1	Top of the lung volume segmented using region growing	81 .
Figure 6.2	Bottom of the lung volume segmented using region growing	81
Figure 6.3	Segmentation of a fragmented lung using region growing	82
Figure 6.4	A comparison of LDM and MLT extraction of ROIs	84

a constant gast matter and the constant and the en mar ing ing ing ing the sould make sould be section in the sector dependence of the dependence of the sector of the · 如果的人们是是这种事实的思想的意思的事情。在4月20日,1月1日,1月1日 ra h- Al Guigiris na suidheagrach adh 70 casirl A. Bh earr r and the set of sets as the speed with the deleters of 网络小鼠藻科学校 网络小鼠属小鼠属美国金属 那些美国子 化合金 1 marsh - Ball marsh and a sharp for the state a shina a that a san a shekar shekar ta barat 1.23 a second and the second se 医马克德氏静脉的膜系上的病 医单位的复数形式 医输出 计可能计可 医克勒克氏神经管 ¢. 网络哈姆 计按照设计过程的 化分离 机分离 机合金运用 and the property of the property of the second sector of the sector of t Station & encoding of stational and encoder of the Astro-2. 增快起路的运行的公司的公式。 gen gen gegen van maak vergengen at anderen 'n trep men wij t 都在这个人,还是这次的现在我的情况,我的问题,不是是一个 a shi ta a alterproperty for the transmission b in the (γ^{λ}) ·第三册 机丁烯 化过去分配 化乙酰乙酸乙二乙酯

List of Abbreviations

 $[1,\infty,\infty]$

CAD Computer Aided Diagnosis CDF Cumulative Density Function CT Computed Tomography DICOM Digital Imaging and Communications in Medicine FP False Positive HU Hounsfeld Units LDA Linear Discriminant Analysis LDM Local Density Maximum MLT Multi-Level Thresholding ROI Region of Interest

11

្រំខ

Chapter 1 Introduction

ang sa sa Bang 👌

Computers have revolutionized the modern world. They have accelerated the flow, presentation and processing of data. In virtually all professions they play a significant role either directly or indirectly. In direct use they can be used as a tool to help make a task more complete, accurate, simple or quick. One realm where computers play a significant role is image processing [1], [2].

Image processing applications have led to computers checking for manufacturing defects during the production process to reading license plates [3]. A particularly important application is medical imaging [4]. Here, computers are used to help diagnose and interpret medical images of patients. Systems have been designed to analyze and interpret things from brain scans to radiograms [5-7]. Technologies such as computerized tomography (CT) and magnetic resonance imaging (MRI) in conjunction with computer hardware advancements have lead to an explosion of research in medical imaging applications.

1.1 CT imaging

CT imaging was invented in the early 1970's by Sir Godfrey N. Hounsfield and Professor Allan M. Cormack independently of each other [3]. CT imaging basically consists of an x-ray source with sensors on the opposite side of a concentric circle which is rotated around the circle's z-axis. A patient lies down, contained within this circle, in between the x-ray source and the sensors. Many images are taken sequentially in the direction of the circle's perpendicular axis to obtain a 3D representation of a particular region of the patient. The intensities in the CT images are representative of Hounsfeld Units (HU) which measure the attenuation of the x-rays as they pass through that particular location. CT images are stored in the Digital imaging and Communications in Medicine (DICOM) file format which keeps the data in a lossless state with up to 15 bits of resolution per pixel. Over the years CT imaging has improved by using less radiation and providing thinner slices to produce more accurate data sets.

CT imaging has allowed doctors to get more detailed and informative images than were available with conventional radiograms. For example, the image in figure 1.1 represents a single image from a 65 slice data set of a patient's lungs where in the past one or two radiogram images would have been taken from the front and sides of a patient in the diagnosis procedure. Slices are cross sections of the patient at specific points, usually a series consists of spatially sequential slices.



Figure 1.1: A CT image of a section of the thorax



1.1

Figure 1.2: a single radiogram taken from in front of a patient representing an entire lung volume

The evolution of CT, and its resultant more detailed images, has given rise to certain challenges. Radiologists now must potentially go through in excess of a hundred images to diagnose a single case where in the past they would only have to look through a handful of radiograms [8], [9]. CT imaging provides a huge volume of useful information which makes it that much more challenging and demanding to evaluate a CT data set thoroughly. Radiologists can often miss important details due to the amount of detail present in a CT data set [10-13]. In this case there is an abundant amount of information but difficulty in processing it. This makes processing CT image data sets and appropriate task for Computer Aided Diagnosis (CAD) systems [12-13]. CAD systems allow for complete and thorough analysis of all the data which can then be used as an aid to the radiologist in their diagnosis. Cancer, for example, is one disease that is commonly screened for using CT imaging.

1.2 Cancer Statistics

In looking at statistics from the US, cancer is the leading killer of people under the age of 85 [14]. In particular it is the second leading killer of children aged 1 to 14 only behind accidents. The life time probability of getting cancer for a man is 46% where for a woman it is 38%. It is also estimated that just over half a million Americans will die from cancer in 2005.

Cancer	Number of expected new			Percentage of expected		
Type		cases		С	ancer case	S
TJPC	men	women	total	men	women	total
Prostate	232,090	0	232,090	32.69%	0.00%	16.90%
Breast	1,690	211,240	212,930	0.24%	31.87%	15.51%
Lung and Bronchus	93,010	79,560	172,570	13.10%	12.00%	12.57%
Colon and Rectum	71,820	73,470	145,290	10.11%	11.08%	10.58%
Pancreas	16,100	16,080	32,180	2.27%	2.43%	2.34%
Leukemia	19,640	15,170	34,810	2.77%	2.29%	2.54%
Total	710,040	662,870	1,372,910			· · · · · · · · · · · · · · · · · · ·

Table 1.1: 2005 US statistics on expected new cancer cases [14]

4

And the second secon

Cancer Type	Number	Number of expected deaths			Percentage of expected cancer deaths		
	men	women	total	men	women	total	
Prostate	30,350	0	30,350	10.28%	0.00%	5.32%	
Breast	460	40,410	40,870	0.16%	14.69%	7.17%	
Lung and Bronchus	90,490	73,020	163,510	30.65%	26.55%	28.67%	
Colon and Rectum	28,540	27,750	56,290	9.67%	10.09%	9.87%	
Pancreas	15,820	15,980	31,800	5.36%	5.81%	5.58%	
Leukemia	12,540	10,030	22,570	4.25%	3.65%	3.96%	
Total	295,280	275,000	570,280				

Table 1.2: 2005 US statistics on expected cancer deaths [14]

Cancer	All Stages %	Local %	Regional %	Distant %
Туре				
Prostate	99.3%	100.0%	100.0%	33.5%
Breast	87.7%	97.5%	80.4%	25.5%
Lung and	15.2%	49.4%	16.1%	2.1%
Bronchus				1
Colon	63.4%	89.9%	67.3%	9.6%
and				
Rectum				
Pancreas	4.4%	15.2%	6.8%	1.8%

Table 1.3: 2005 US statistics on 5 year survival rates [14] depending on what stage the cancer is found. Local means the cancer has not spread, Regional involves the cancer spreading to lymph nodes or organs in the immediate region and Distant refers to cancer which has spread to parts of the body that are remote from the origin of the cancer.

Table 1.1 shows the expected deaths and new cases of some of the most common forms of cancer. Men and women are most susceptible to prostate and breast cancer respectively [14], [15], but the survival rates for both are excellent when caught early. Lung cancer is the second most likely form of cancer for both men and women and the third most common over all. In table 1.2, however, it can be seen that lung cancer accounts for the most cancer deaths of any cancer; it actually accounts for more cancer deaths than leukemia, prostate, breast, colon and rectum cancer combined. As seen in table 1.3 a lung cancer patient has a significantly better chance of survival if they are diagnosed early. Based on the overall survival rate for lung cancer it can be observed that the majority of lung cancer is detected after it has spread significantly. If the survival rate could be brought closer to the early detection rate for lung cancer then a truly significant number of lives could be saved. Assuming good detection techniques could bring the overall survival rate up to 40% then that would equate to approximately 42,000 lives saved each year due to early detection in the United States alone.

1.3 Lung Nodules

It is because of the potential to save lives and continued advancements in computing power that automated lung nodule detection has become an active research area in the medical imaging community for CAD systems. Lung nodules refer to growths within the lungs that are potentially cancerous. Nodules are generally defined to be any growth in the lungs that is smaller than a 30 mm sphere [8], [9], [16], [17]. Figures 1.3 through 1.7 give good illustrations of some of the different types of nodules that can be found in a CT image.

1.5



122

133

Figure 1.3: Circled on the left side is a typical lung nodule; just below two more nodules are circled which are touching the lung wall (juxtapleural nodules)



Figure 1.4: Near the middle of the lung a ground glass nodule is circled and just below it is a nodule in contact with the lung wall



Figure 1.5: A nodule is circled that is overlapping with vascular structures, in this case a blood vessel at the bottom right portion of the nodule



Figure 1.6: A pair of nodules overlapping with the diaphragm as it enters the CT data set near the bottom of the lungs



Figure 1.7: A large nodule attached to the lung wall near the top of the image is circled, as well there are many other nodules in the image that are left unlabeled

Nodules are generally sub-classified as solitary nodules (fig. 1.3), juxtapleural nodules (fig. 1.3, 1.4, 1.7) and ground glass opacity nodules (fig. 1.4). Solitary nodules are generally well formed isolated nodules, which tend to be circular but can also have a strongly elliptical shape. Juxtapleural nodules are nodules that are in contact with the lung walls. Their shape tends to be less defined than isolated nodules and they do tend to get bigger than isolated nodules, though they can also be very small. Ground glass opacity nodules are just basically areas of opaque irregularities that do not tend to be particularly well formed. Ground glass nodules often have contact with the lung walls. All of the different types of nodules can overlap with normal vascular objects; an example of this can be seen with a nodule overlapping a vessel in fig. 1.5. In fig. 1.6 two nodules are seen that are overlapping with the diaphragm as it enters the CT data set from below. A very large nodule is found in fig. 1.7 which reflects the tendency for some juxtapleural nodules to be very elliptical. The nodule in this case is growing along the contour of the lung where some juxtapleural nodules may be attached by only a faint connection as can be seen in one of the nodule in fig. 1.3. One of the main points to take

from this description is that nodules do have tendencies based on their respective classifications, even with these tendencies there are always exceptions in practice. Intensity, shape definition, location and connectivity are all variables that provide for many different types of contexts and appearances given nodules of the same basic type. This makes lung nodule detection a challenging and necessarily robust process for a computer system.

1.4 CAD System

Lung nodule detection consists basically of three steps. First, the lung volumes need to be segmented from the body. Though this may seem to be a simple and straightforward step it is a very important one that does demand careful consideration. Next, Regions of Interest (ROIs) are extracted from the lung volume to include all objects that are remotely suspicious of being cancerous. This step is usually closely integrated with the final step so they can sometimes be grouped together as a single stage depending on the approach used. The final step is False Positive (FP) reduction. This step takes all ROIs and tries to eliminate as many candidates that are not nodules while retaining as many suspicious candidates as possible. This last step is key to delivering an effective CAD system because it greatly affects the confidence behind the objects which are labeled as potential nodules.

A complete lung nodule detection system that builds upon effective components from previous approaches and uses original techniques to deliver a unique and innovative system has been developed. It contains a new approach in lung segmentation that uses 3D region growing to provide more complete lung volumes by including parts of the lungs at the top and bottom of the CT data set that are eliminated with traditional volumetric approaches [18]. An adaptive, non-linear multi-level thresholding (MLT) approach is used that evolved from an existing fixed, linear MLT technique for ROI extraction [8], [19]. The adaptive and non-linear aspects use the cumulative density function (CDF) of the lung volume to make the approach more robust. The FP reduction stage uses original and creative approaches such as a "tight" bounding box metric, vessel and lung wall connectivity, 3D compacting and analysis and vertical contrast to produce the final output of the system.

The effectiveness of this work along with the results obtained by other researchers shows significant promise for using a CAD system in the area of lung nodule detection to increase the accuracy of diagnosis by taking advantage of the information available through advances in CT imaging [17].

1.5 Outline

The next chapter contains the literature review. It presents different approaches that have been presented for lung nodule detection. The merits and drawbacks of each approach are analyzed, and based on these a course of action is chosen for the development of a nodule detection system.

Chapter 3 deals with a lung nodule synthesis technique that was developed as a precursor to the actual lung nodule detection system. The purpose behind the technique along with its value within lung nodule detection is discussed. It is used to create a preliminary test set for the nodule detection system, and the results of the system's performance on this test set is also presented in the results section, chapter 6.

Chapters 4 and 5 are the focus of the research and present the entire design of the lung nodule detection system. Chapter 4 explains the lung segmentation procedure, while chapter 5 deals with ROI extraction and FP reduction. These chapters contain the methodologies and the parameters under which the system performs in its final state of testing.

Chapter 6 provides all of the results for the system at different stages of development along with the changes the system underwent as it progressed in the experimentation process. Critical analysis at each stage is offered, and through this analysis the changes to the system are conceived and implemented.

The last chapter, chapter 7, deals with the conclusions that have been made about the system at its current stage of development. It also points out where the system needs improvement, along with offering many avenues that can be pursued towards this end. and the main and a manual frequency of the construction of the state of the

Chapter 2 Literature Survey

Automated lung nodule detection is a relatively new area of research. One of the first papers in the field of lung nodule detection using CT images was published by [20] based out of the Department of Radiology at the University of Chicago. This early work helped to set in motion future research in the area, especially as computers got faster and CT imaging became more prevalent. It was only a matter of time before technology caught up enough to be able to design effective CAD systems. Now, a basic home PC can be used for image processing work that would have been unmanageable for most powerful computers back in the mid-1990s.

Now, research in this area is increasing and CAD systems are showing promising results. The results and approaches of more developed techniques such as the ones by [8] and [9] using Linear Discriminant Analysis (LDA) and rule based approaches are used as important input into more recent approaches. Recent work by [17] is an excellent example of evolving and refining the techniques of previous work while adding on their own unique contributions. Others look to apply existing computer image processing techniques such as template matching as done by [21] and neural networks as presented by [22]. The number of approaches available allows researchers to examine their effectiveness and choose their own path for developing a lung nodule detection technique.

2.1 LDA Based Techniques

A system using LDA tries to exploit statistical feature differences between different classes to separate them. The system developed by [8] uses thresholding and LDA. The body is first separated from the background by analyzing the grey level values of a diagonal line from the top right corner to the center of each CT image. A threshold value is obtained from an analysis of this line and an outer thoracic border is obtained to separate the background from the body pixels. To segment the lungs a grey level histogram of the body pixels is analyzed to obtain a threshold that provides a maximum distance between pixels representing body and non-body pixels. A contour of the lungs is obtained and used as the initial lung border. To fill in any holes the lung contour is processed by a rolling ball filter to smooth the contours.

For the ROI stage they apply multiple grey-level thresholding on the lung volumes. At each level, all objects that exist above the threshold are considered to be nodule candidates if their volume is less than that of a 3 cm diameter sphere. The thresholds are evenly spaced out and are at fixed levels for all cases.

The final stage of the process involves FP reduction. The system computes nine nodule features: volume, sphericity, radius of the equivalent sphere, maximum compactness, maximum circularity, maximum eccentricity, mean grey level within the nodule, the nodule's grey level standard deviation and the threshold at which the object first decreases below the volume criterion. These features are then put into a LDA classifier to produce the final nodule candidates.

This approach by [9] provides some positive tactics. It uses multi-level thresholding (MLT) which is a good way to try to separate objects within the lungs based on their intensity differences. The volume restriction tries to ensure that nodules are not assimilated into larger objects. The features used for LDA focus strongly on the candidates shape and intensity characteristic, while only using one indirect feature (the threshold at which the object first decreases below the volume criterion). One of the deficiencies in the approach is the lack of nodule context. The features focus almost exclusively on grey level content of the nodule and the geometric properties without having any direct way to evaluate the context of the nodules within the lung volume. In addition, the thresholding approach uses fixed threshold levels and spacing. This will be problematic if the CT scans analyzed are darker or brighter as a whole than what is expected by the pre-selected thresholding levels. By fixing the thresholding levels the system can not adapt to variations in the distributions of the CT scans.

The reported results of their early work [8] list a sensitivity of 70% and approximately 3 FPs per CT slice. Future work using the same features and systemic approach drew FPs down to approximately 1.5 while retaining 70% sensitivity [19]. They improved their system by adding rule based elimination to the FP reduction stage.

. t.

2.2 Rule Based Techniques

Rule based techniques try to detect nodules using a set of features which are then applied to binary rules, the outcome of which are used to identify nodules. One such technique was developed by [9] and focuses heavily on rule based feature analysis with **a** strong emphasis on nodule context.

The first step involves extracting the lung fields from the CT data set. This approach involves using a fixed threshold to perform an initial segmentation of the lungs. To only include the lungs in the segmentation other objects that would also fall below the threshold such as the trachea are eliminated based on their cross-sectional and intra-series positions. The remaining volumes are considered to be the lungs. Contours are taken for the lungs on each slice where they exist. The contours are then smoothed to fill any gaps along the borders. This approach is similar to the one performed by [8] but it uses border pixel analysis along with shortest point linear connections to fill gaps. The effect is very similar to the rolling ball algorithm mentioned earlier. The lung volume is then analyzed and split up into 4 sections where groupings of adjacent CT slices belong to a particular section.

Before extracting the ROIs the lungs would be preprocessed. The lungs would be smoothed by using basic morphology operations followed by unsharp masking to improve ROI extraction by providing more strongly defined and separated structures. After the preprocessing the ROI extraction phase begins.

The lung regions are separated into air and organ clusters. The nodule candidates belong to the organ cluster. They then apply thresholding on the organ cluster to extract the ROIs. Next, they perform surface curvature analysis in the intensity plane of the ROI candidates in order to further refine them. With the ROIs defined they now proceed to rule based FP reduction.

The first stage of FP reduction involves extracting the feature set for each ROI. The features are: area, thickness, circularity, grey-level, variance of grey-level, localization, variance of gradient, distance from the lung wall, convexness and contrast. From these features they develop rules which deal with nodules in different cases. They take separate approaches for nodules which are interior and those which contact the lung wall.

One of the major positives about this approach is distinguishing interior nodules from lung wall nodules and evaluating them using separate rules. Another good point to note is that the features chosen do make an attempt to take the context into account. The convexness and the contrast features are used to analyze the surrounding pixels for nodules contacting the lung wall.

The reported results for this system were 90% sensitivity and 0.3 FPs per slice. These good results along with a well thought through approach make this a very promising system.

Another rule based approach was developed by [23]. For lung extraction a grey level histogram of the data set is analyzed and a threshold, to separate the lungs from their surroundings, is found based upon separating the two major peaks of the histogram. Volumetric analysis and 3D connectivity is used to form a complete lung volume. Morphological closing is applied to fill in any gaps and to smooth out the lung contours.

For ROI extraction a 3D local density maximum (LDM) algorithm is used. The lungs are thresholded from an intitial value to a bottom value incrementally in evenly spaced steps. The top threshold can be chosen as the highest pixel value in the lung volume while the lowest can be the smallest. Objects are identified at each threshold and, if taken as a ROI, are kept separate by not allowing them to be assimilated into larger objects at lower thresholds. This attempts to separate connected components with different density profiles.

A rule based approach is used to reduce FPs using three features: object volume to modified bounding box volume ratio, maximum depth to cross-sectional projection length and major to minor cross sectional bounding box ratio. These features focus strongly on shape properties while also looking at the compactness of the ROI.

The results produced a sensitivity of 84.2% with 5 FPs per CT case. It should be strongly noted that these results were performed on a set of 266 simulated nodules as the researchers did not have access to a real CT database of lung nodules. This technique presents a new approach to ROI detection that is geared towards being able to separate connected components. This is an important goal because nodules due tend to overlap with other structures which makes them more difficult to segment properly. Also, the rule based approach is a simple approach that focuses on 3D structure. This technique was implemented for this thesis because of its simplicity and good reported results. Experiments found the system to perform very poorly on our CT data; further details are presented in the results section, chapter 6.

2.3 Template Matching Techniques

A template matching technique attempts to use a prior database of nodules that can be used to find similar objects in the target analysis. Rules are often used after the template matching step to reduce the FPs.

A template matching approach is attempted by [21]. This system is based on matching nodule candidates to members in a pre-existing nodule template database while using rules to reduce the number of FPs.

The basic premise of template matching is that a database of nodule templates could be used to find similar candidates in the lung volume. They used a spherical model with a Gaussian distribution based on their analysis of the characteristics of real lung nodules. They used four 3D lung nodules which were evenly spaced in diameter from 10 to 40 pixels, where one pixel was 0.638 mm. To detect nodules that appear in only one slice, the middle sections of each 3D model were used as 2D models. A genetic algorithm was used to find nodule candidates by locating areas that were similar to any of the members of the template library. They also performed a separate search along the lung walls for ROIs that did not use a genetic algorithm. They would use a general template matching technique by going around the border of the lungs and using a modified version of the template library. The template library was modified by stretching the Gaussian models in the x-direction and using halves of those ellipsoids as the new templates. For

the lung walls the templates were rotated to be tangent with the lung wall surface for each point of inspection.

For FP reduction they used a rule based approach incorporating a set of 16 features: mean, standard deviation, area, circularity, irregularity, contrast, max mean CT value, directional variance of pixel gradient, directional cross-correlation of pixel gradient, inverse difference moment, entropy, second area, second mean, local mean, local standard deviation and local directional variance of pixel gradient.

Their results for the technique indicated a sensitivity of approximately 72% with 5.5 FPs per CT case. This was an improvement over their previous work [24] where they had the same sensitivity but a much higher number of FPs at 30.8 per CT case. In the previous experiment they used only 11 features as opposed to 16 in the current one for FP reduction.

One of the major challenges with this kind of approach is the nodule template library. In this case the library was relatively small but even with a more comprehensive library it remains difficult to try to account for all possible nodule cases. One of the major positive contributions from this paper is the realized FP reduction by improving a rule based technique from their previous approach.

2.4 Fuzzy Neural Network Technique

This approach uses a combination of fuzzy logic with traditional neural network based training. This approach is interesting as it tries to blend together two approaches which separately can work well for feature based decision problems.

A fuzzy neural network approach was presented by [22]. To segment the lungs they use a threshold value to form a binary image. This image is then median filtered and operated on by morphological closing. Finally, region growing is applied to fill any gaps remaining in the lung fields.

For the ROI acquisition mechanism a histogram analysis is performed. The grey level value that retains the top 20% of the pixels in the lung volume is used to threshold the image. The objects that remain after the tresholding are the ROIs. For feature reduction, a fuzzy neural network approach is used. As inputs to the network three features are measured for each ROI: area, circularity and mean grey level. A fuzzy neural network is a combination of fuzzy logic and neural networks. The main premise behind fuzzy logic is that it is an intuitive way of looking at problems, instead of using specific equations general truths are used, implemented through rules [25]. These rules provide the appropriate guidelines for a solution. For example, in this implementation the area feature is assigned a membership value between 0 and 1 for three fuzzy membership functions that determine the ROIs membership to three different categories of size: small, medium and large. All of the membership functions are then applied to rules and then into a defuzzification layer which determines whether the ROI is a nodule or not. The neural network portion of the system is responsible for the weights assigned to the inputs going into the defuzzifaction layer. The neural network uses the gradient steepest descent method [22] to train using a set of 20 patterns (12 nodules, 8 non-nodules) cyclically until the mean square error falls below a certain target.

Their reported results were a sensitivity of 89.3% with 0.3 FPs per CT slice. The target size of objects they can detect is between the areas of circles with diameters in the range of 10 mm to 50 mm.

Some of the positives from this approach are that fuzzy logic is a potentially good candidate as an approach for lung nodule detection. Fuzzy logic is very effective in techniques that need more than a binary process [25–27]. Applying a neural network to a fuzzy system is a good attempt at integrating two systems that are often used for problems that depend on feature analysis. The implementation of the system appears to need more refinement. The ROI acquisition stage is a very simple single level thresholding approach, which is a regression from a MLT approach in terms of thoroughness. The number of features used is limited to three while providing no context for the ROI in terms of 3D features or its surroundings. Also, the neural fuzzy network is trained using a very small set that consists of only 20 unique patterns which is not ideal for a robust problem such as lung nodule detection. Also, the target size of nodules being over 10 mm limits the effectiveness of this approach as a significant portion of nodules tend to be below 10 mm in size. Though this approach does introduce an interesting

fusion of two potentially effective techniques the implementation and execution need to be refined to explore the true potential of this kind of a system.

2.5 Hybrid Rule Based and LDA systems

This kind of approach tries to combine rule based techniques and LDA based techniques. Both have been used as the primary form of FP reduction separately but in this kind of approach both techniques are fully implemented into the system.

A combination of a thorough rule based approach and LDA was proposed by [17]. For lung segmentation a fixed low threshold was used to segment the body form the background initially. The lungs were extracted using k-means clustering. In this case k = 2 since there are only two classes, air and body pixels. The lungs were further separated into sections by splitting lungs on each CT slice into a peripheral sub-region and a central sub-region which would be used later for FP reduction. To fill discontinuities along the lung borders they used indentation detection [17] which similarly to previous techniques used a straight line to connect gaps in the lung borders.

To obtain the ROIs a k-means clustering technique was used with k = 2, similar to the approach used for segmenting the lungs. The number of pixels belonging to each class was related by a ratio which varied depending on where the clustering was being done. For upper slices and lower slices less ROIs were expected so the algorithm used a higher ratio of background to object pixels. For the middle portion of the lung the ratio was smaller so more objects could be detected. Some of the outputs due to this technique would contain holes; these were filled since nodules are considered to be full structures.

FP reduction was a multi-stage process. First 2D rule based reduction was applied, based on the ratio of major to minor axis of an equivalent ellipse followed by a compactness criterion. The remaining 2D objects would then be connected using 26 point connectivity. Four more rules were then applied for these objects relating to the bounding box size, maximum circularity and a relation of the location of the object to its size. At this point 3D features were extracted from the remaining candidates: volume, surface area, mean grey level, grey level standard deviation, skew and kurtosis of the grey level histogram. The last part of FP reduction involved applying LDA. For the LDA process

the following features were used: volume, surface area, mean grey level, standard deviation of the grey levels, maximum cross sectional area, perimeter, major and minor axis, circularity and eccentricity.

The results after the rule based elimination were a sensitivity of 87% and 2.92 FPs per slice. After applying LDA the sensitivity fell to 84% while reducing FPs to 1.74 per CT slice.

This approach results in a thorough and promising combination of LDA and FP reduction. The approach uses an extensive and effective combination of rules and features to perform FP reduction. The rule set was large and tried to account for nodule context and internal properties. The LDA performance suggests it can be an effective tool to use in order to reduce FPs while retaining good sensitivity performance.

2.6 Automatic Lung Segmentation System

A system focused solely on automatic lung segmentation was produced by [18]. In most papers lung segmentation is treated as a precursor to nodule segmentation, however, in this paper lung segmentation is the focus and is addressed with more depth and attention. This is beneficial because proper lung segmentation is very important in the proper identification of nodules along the lung walls as well as producing complete and proper lung volumes.

The process begins with choosing a threshold for the lungs. The threshold is based on a technique called optimal thresholding [18]. This tries to find a threshold that will separate air voxels from body voxels through an iterative feedback process. Once the threshold value is found it is applied to the CT scans the background is eliminated by finding air voxels that are connected to the borders of the CT images. The lungs are obtained by only obtaining those air groupings which meet a minimum size criterion. After this volumetric based reduction the trachea is identified and eliminated using a slice by slice region growing approach. Topological analysis is used to fill in any enclosed holes left in the lungs after thresholding. To fill in any indentations left along the lung walls a morphological closing is applied. This paper presents thorough and detailed guidelines for lung segmentation that are not present in works where the main focus is lung nodule detection. They also provide a performance metric by comparing their segmentation results to manual traces of the lung regions. The mean difference between the borders of the manual traces and the lung segmentation system was 0.48 pixels.

2.7 Comparing Results

Comparing results to objectively evaluate the relative performance of different techniques is very important. For proper comparison you would have a robust, publicly available test set that all techniques could use. This way the results would have merit relative to each other and also in an absolute sense due to the integrity of the test data. Currently there are no such databases but there is a government initiative in the US to produce one [28]. This initiative has produced a preliminary 23 nodule database that is intended to contain 400 nodules when it is complete.

Given the current climate, all the researchers to date have needed to find their own data sets and obtain results from them. This point especially makes it difficult to compare techniques based on their reported results since there is no absolute way to judge the differences between nodule sets. In addition this makes it difficult to implement learning algorithms in a system because researchers are limited with the cases they have access to. As alluded to in the introduction there are many different types of nodules that can exist in many different contexts. Even with access to a significant amount of data it would be difficult to gauge if it was complete enough to produce a completely robust system.

The problem of objectively comparing the results of differing techniques is one that has been recognized and investigated to some degree [17]. An example of the disparity that can occur for the same system using different data sets was examined by [17]. This problem is further addressed in the results section, chapter 6, which includes an evaluation of the technique produced by [23].

The results in table 2.1 can show how much the data set can affect the performance of a system. This is reflected very strongly in the system by [29] which had

excellent results on their initial data but failed significantly on a subsequent data set [17]. The system in [8] had weaker initial performance but their approach proved to be more robust and they actually improved their FP performance by refining their system [19]. The main point that is reinforced in table 2.1 is that the experimental results can not be used as an absolute comparison metric since everyone is using different test data. From table 2.1 it can be seen that the technique that had excellent results initially failed on a different data set.

System	Initial Sensitivity	Secondary Assessment	Variation	Initial FPs	Secondary Assessment	Variation
Fiebich et al.	95.7%	30%	-65.7%	0.3	6.3	6
Armato et al.	72%	71%	-1%	4.6	1.5	-3.1

 Table 2.1: Performance results of nodule detection system in initial and follow up studies

 from [17]

Many of the techniques listed involve a rule based stage for FP reduction. A rule based approach is an excellent foundation for a nodule detection system as it can be easily controlled and it is easy to directly implement advice from radiologists. The work by [9] is entirely rule based and has good reported results. The reported performance results for the techniques that have been listed are in the range of 70% to 90% where FPs have ranged between 0.5 to 4.6 FPs per slice. These are good general guidelines for the working parameters of a system since this encompasses a wide range of performance results using different systems over different data sets.

The objective is to make an initial attempt at creating a robust system, but more importantly one that can serve as a foundation which can be built upon. For this reason it was initially desired to design a simple rule based approach inspired by the work of [23] who have delivered a system that uses a simple rule based FP reduction stage along with an innovative new ROI extraction algorithm. This approach is later examined in the
results section, chapter 6, and is found to have significant problems. For the development of the lung segmentation stage most papers do not produce much detail so the work done by [18] served as a good model from which to develop a lung segmentation technique.

en 1997 var en der der Bereik der Strick gestigt stellte sonder der einen einen seinen son der einen son der ei Stellter son der Stellter von der Stellter stellter stellter stellter stellter stellter stellter stellter stellt

Chapter 3 Lung Nodule Synthesis

The most important resource needed for developing a lung nodule detection system is a lung nodule database. Unfortunately, there have been no public data sets until recently [28] so obtaining test data has been a challenge in and of itself. Even this data set is preliminary and is still being completed. To obtain data researchers partnered with hospitals to gain access to the resources they need. For schools that can not establish these partnerships research in the area of lung cancer detection has been difficult. To work around this problem some researchers have created data sets by creating synthetic nodules which they then insert into the test data [23], [30], [31]. Synthetic nodules can even be used to test techniques built on real nodules or to potentially increase a limited data set that might be available to a researcher. This can allow the researcher to artificially create cases that are not present in their data or to create an artificially large database to evaluate the potential of using synthetic nodules as input into learning based approaches. Corner cases are an excellent application of a lung nodule synthesis technique as even with a real nodule database it can be difficult to find them. Learning based techniques usually require a large and diverse set of inputs to train the system to a point of stability; synthetic nodules have the potential to produce the volume and context necessary in such a data set. In these regards synthetic nodules can be useful for a variety of development and testing purposes that are limited by a clinical CT data set.

3.1 Lung Nodule Synthesis Approach

CT images of real lung nodule cases formed the basis for the lung nodule creation technique [30-38]. The primary focus of this work was to create nodules which are

relatively circular. This allows the nodule synthesis system to target a particular structure in order to clearly define the goal of the synthesis technique. As already discussed, nodule shape and size can vary significantly, so it is important to have a clear focus for the type of structure to be simulated. Using a circular model, which allows for elliptical deviation, permits the system to simulate a large subset of nodules. There is freedom to set the context of these nodules as they can be placed anywhere in the lung volume.

In figure 3.3 the nodule resembles a sharp bell shaped structure. Based on this observation a Gaussian shape was used for the synthetic nodule structure. A Gaussian distribution has been used as a structural base in other experiments involving synthetic nodules [23]. It has also been used as the foundation for nodule templates in detection techniques [24]. The magnitude and standard deviation of the Gaussian are all parameters that can be manipulated. The following formulas are used to create the Gaussian that forms the basis for the synthetic nodule [3]:

$$G(x, y) = e^{\frac{-(x^2 + y^2)}{2\sigma^2}}$$
(3.1)

The x and y terms in (3.1) refer to spatial coordinates along the x and y axes in a 2D surface and the σ term refers to the standard deviation of the Gaussian. In (3.1) the scaling term is omitted in order for the Gaussian G(x,y) to be equal to one at its center. This result, G(x,y), is then sampled to create an M×N matrix, G[m,n], centered on the origin. The samples are taken using Δx and Δy equal to 1 as the sampling period. The Gaussian that is created is a perfectly symmetric 2D intensity image; however, the original nodule is clearly not symmetric as seen in figure 3.2.

and a second second

All of the second states of the second states of the second states and the second states are second states.



Figure 3.1: A CT scan containing a nodule in the right lung (left side of image)



Figure 3.2: The image from figure 3.1 zoomed into the lung nodule which has been circled



Figure 3.3: Mesh of real lung nodule from Figure 3.2

The next step in the creation of the synthetic lung nodule addresses this issue. A random, asymmetric matrix is added to the Gaussian to provide it with an authentic look that would not be present if the image was a perfectly symmetric Gaussian.

$$R(m,n) = random_{m,n}$$

$$random_{m,n} \in [-1,1], distributed uniformly$$

$$F(m,n) = \alpha \times G[m,n] + \beta \times R[m,n]$$
(3.2)

In (3.2) α and β are scaling terms. These are used to control the magnitude and asymmetry of the nodule during the nodule insertion process.

The final step is to take the modified Gaussian, F[m,n], and insert it into a real CT image. This is done using a Graphical User Interface (GUI) that reads in the DICOM image data. To blend the nodule smoothly into the image any values of the nodule that are less than those of the image at the insertion coordinates are replaced with the original image value.

The parameters that determine the appearance of the nodule are accessible in the GUI. These parameters are the peak intensity and standard deviation of the base Gaussian, the dimensions M and N of the sampling matrix and the interval of the random matrix. The process of inserting the nodule involves adjusting the parameters until the inserted nodule appears as desired by the user. From other studies real nodules were observed to have some common characteristics. Nodule diameter is generally less than 30 mm, and they tend to be circular and slightly elliptical. For one nodule detection technique small synthetic nodules with a diameter between 2 and 7 mm were created [23]. These nodules were restricted to circular shapes or ellipses where one diameter did not exceed the other by more than a factor of 1.5. A different study [17] that used a database of 64 real lung nodules provided statistics of their diameter values. Figure 3.4 contains the diameters of the nodules in this study, where a value of 8 mm means the diameter was in the range (4 mm, 8 mm]. The diameters had a mean of 8.89 mm, standard deviation of 5.66 mm, minimum of 2 mm and a maximum of 25 mm. About 78% of the nodules in their database had a diameter less than or equal to 12 mm.



Figure 3.4: Distribution of nodule diameters observed in the test data used by [17]

To produce credible lung nodule images the feedback of a radiologist, or someone who has professional experience with lung nodules in CT images, is desired. This is important to produce nodules with a realistic density distribution, shape and placement since visual assessment is the primary mode of verifying the realism of the inserted nodules. Ideally, the person using the GUI to create synthetic lung nodules would be an experienced radiologist.

3.2 Lung Nodule Synthesis Results

An example of a synthetic nodule is presented in figure 3.5 alongside a real nodule.



Figure 3.5: A real nodule and a synthetic nodule below it are identified by a circle

From figure 3.5 it can be seen that the two circled objects are very similar in terms of intensity and size. Also, the synthetic nodule is more circular than the real nodule in this particular case though this is adjustable. Looking at the synthetic nodule asymmetry is clearly present in the nodule which provides needed authenticity to the nodule.

A radiologist confirmed that a set of 23 synthetic nodules that were created had an authentic visual appearance. In general, nodules are judged primarily using visual analysis by a radiologist so the quality of the synthetic nodules was validated after they passed visual inspection by an expert radiologist. Inspection of the mesh plot in figure 3.6 for the synthetic nodule also yields **a** good comparison to the real nodule depicted in figure 3.3. The shapes are similar to each other. The synthetic nodule does have the fundamental Gaussian shape as expected from (3.1) and the asymmetry gives it a gentle, natural noisy appearance in both the mesh and intensity images. The nodule does blend very well with its environment as seen in figure 3.5. The blending approach of using the largest pixel value between the nodule and the background is used instead of filtering as it tries to alter as few pixels in the image as possible. Sharp drop offs for very dark backgrounds are not a concern as nodules can have very defined borders and high contrast with their immediate surroundings.



Figure 3.6: A mesh of the synthetic nodule from figure 3.5

The value behind creating synthetic lung nodules was three fold. Initially, there was no available CT data set with lung nodules so this was going to provide, at the very least, some data to experiment on. Secondly, it provided an opportunity to learn more about nodules and their properties through studying them more extensively. Lastly, it provided the opportunity to expand our own CT data primarily for testing. Inserting the nodules as irregular separate structures should cause them to be detected regardless of

their exact density distribution due to the simple fact that they are abnormalities within the lung volume.



Figure 3.7: A screenshot of the nodule synthesis GUI

Chapter 4 Lung Volume Segmentation

Lung volume segmentation is a necessary step in a lung nodule detection system. Lung nodules are found within the lungs so separating them from the rest of the CT data is important in order to target a lung nodule detection system only in the areas where lung nodules can occur. In addition to being a necessary step, it is also the first stage of a lung nodule detection system, the basic outline of which is shown in figure 4.1.



Figure 4.1: General structure of a lung nodule detection system

There have been few publications that deal in depth with lung segmentation [18], [39] which is unfortunate since good lung segmentation is an important factor in the total

process. By ensuring the lung volume is completely segmented the best opportunity is afforded to find all possible lung nodules. A simple and effective approach was provided by [18].

4.1 Lung Volume Thresholding

The initial step in segmenting the lungs is to find an appropriate threshold to begin with. To find the threshold value a modified version of the technique used by [18] is used. The CT values are considered to consist of body pixels and non-body pixels, finding a threshold attempts to identify a good boundary between these two regions.

In figures 4.2 and 4.3 a single CT slice is presented and analyzed. Figure 4.2 (b) shows the area of a CT image that represents the actual scan. This area is intuitively correct since CT scans occur in tube shaped chambers which lead to circular CT scan areas. Figure 4.3 shows the HU distribution of the CT image in figure 4.2 giving an idea of the intensity value distribution of the slice. In general, the area between -800 HU and -200 HU in a CT scan of the body tends to be a valley but in this case there is a hill like distribution. This makes finding the appropriate CT threshold value particularly challenging for this slice since lung segmentation techniques usually try to find a value that is directly between the two major peaks at -950 HU and 150 HU which represent non-body and body pixel clusters respectively [18]. This approach is founded on the belief that a shallow valley exists between the two clusters. It is important to note that non-body pixels denote air and low density tissue where body pixels refer to high density objects within the body.

化合金 建制造工作 化乙基乙基乙基乙基乙基乙基乙基乙基乙基乙基乙基乙基乙基乙基乙基乙基乙基



Figure 4.2: (a) a single CT scan. (b) The effective CT scan area of figure 4.2 (a), areas in black are not part of the actual CT data



Figure 4.3: The HU distribution of the CT scan in figure 4.2

The approach by [18] tries to separate the body and the non-body pixel clusters using an iterative approach. The threshold value T_i defines the threshold value at iteration *i*. The initial value is $T_0 = -500$ HU the midpoint between -1000 HU and 0 HU the approximate values for air and water respectively. The mean values of the non-body and body pixels are obtained at each iteration, represented as μ_n and μ_b .

$$T^{i+1} = \frac{\mu_n + \mu_b}{2}$$
(4.1)

The process described by (4.1) continues until T^{i+1} is equal to T^i rounded to two decimal places.

This approach works well in dividing the HU distribution in a manner that provides a large separation between the major peeks of the body and non-body pixels in a CT image. This approach works well for cases where the two major peaks are separated by a shallow valley, as is the case in most CT scans. This does not work well for the CT slice denoted in figure 4.2, the reason being that the two peaks are separated by a hill-like distribution. Clustering into two relatively evenly split sets becomes challenging and ineffective in this case.

The key to resolving this problem is that even in cases where the two peaks are separated by a shallow valley there is a large range of acceptable values that can be chosen for the threshold. The value can fall anywhere in the valley between the bases of the two peaks and the threshold will provide a good separation. This means biasing the threshold to the body cluster peak would be acceptable in the general case and in the case where the shallow valley does not separate the two major peaks in the HU distribution. This system uses the algorithm of (4.1) but it does not include any pixels below -874 HU in its calculations. This value was found experimentally by adding 250 to the HU floor of 1024. This biases the threshold value higher while not being affected by rare distributions since the adjustment is absolute and not density dependant. Using the system's approach the threshold was found to be -220 HU where using all pixel values the threshold was found to be -440 HU.

As can be seen from figure 4.4 using the higher threshold value included a significant portion of the lung volume in the non-body pixels that was not present in the initial thresholding. A significant portion of the lungs in this case would have been omitted without using a threshold value closer to the body pixel cluster in figure 4.3. By using the higher threshold value it can be observed that the lung volumes were segmented properly by the thresholding step.

and an in the second second and the second second









(c)

Figure 4.4: the results of thresholding the CT slice in figure 4.2 with HU values of -440 (a) and -220 (b) where the white pixels are considered body pixels. The image in (c) shows pixels that are reclassified as non-body pixels when comparing (b) to (a).

In figure 4.5 a more normal example of a CT slice containing the lungs is seen. The HU distribution in figure 4.6 shows off a more typical distribution for these kinds of CT images characterized by a shallow valley between the two major pixel clusters. Using the approach by [18] the threshold obtained was -440 HU and by our technique it was -363 HU.







Figure 4.5: A sample CT image with normal contrast and HU distribution



Figure 4.6: The HU distribution of figure 4.5, in this case characterized by a shallow valley between the two major pixel clusters

An example of thresholding a normal CT slice containing the lungs using both approaches is presented in figure 4.7. Both thresholds do a good job of segmenting out the lung volumes. It can be seen that the two segmentation results provide the same basic segmentation by observing the difference image in figure 4.7 (c).





(c)

Figure 4.7: the results of thresholding the CT slice in figure 4.5 with HU values of -440(a) and -363 (b) where the white pixels are considered body pixels. The image in (c) shows pixels that are reclassified as non-body pixels when comparing (b) to (a).

Now that this thresholding approach has been shown to be robust and effective the process can proceed to the next step of actually extracting the 3D lung volume.

4.2 3D Lung Volume Extraction

With a thresholding approach being well defined the rest of the lung volume segmentation process continues. The first step after thresholding involves removing the background pixels. This is done by eliminating 2D 8-connected non-body components that are contacting the image border.



Figure 4.8: The grey pixels around the white pixel are considered connected to the white pixel in an 8-connectivity scheme

Figure 4.16 is the result of eliminating the non-body pixels connected to the border of the image in figure 4.15. Here it is seen that the majority of the background is eliminated save for some small objects around the lungs and an artifact at the bottom of the image which resulted form non-body pixels being isolated from the image border. These artifacts are eliminated later on.



1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 -

Figure 4.9: The results of thresholding figure 4.5 using a CT value of -363 HU where the non-body pixels are represented in white



Figure 4.10: The result of eliminating all non-body pixels which are connected to the image border in figure 4.9

The next step in the process involves removing any holes that are completely contained within a non-body object. This consists of finding body pixel objects which have no contact with the image border. From figure 4.10 many body objects can be seen that are represented with black pixels contained completely within the lung objects. The results of removing these isolated objects can be seen in figure 4.11.

and the provident of the second states of the secon



Figure 4.11: The results of removing the isolated body pixel objects from the lungs in figure 4.10

After removing the holes from the lungs the next step in the process involves finding a seed image for the rest of the lung segmentation process. The seed image is used to perform 3D region growing to allow for a complete and well defined lung volume. The technique developed by [18] for complete lung volume segmentation is used only to find the seed image for the system's lung segmentation process. The existing lung segmentation technique basically retains all non-body objects on the current slice which exceed 1% of the total pixel count of the image. In this case the image is 512×512 pixels meaning the 1% limit is any object above 2621 pixels. If more than two objects exceed this minimal limit than the two largest ones are retained. This criterion effectively eliminates the trachea from the segmentation since it is generally too small to exceed 1% of the total number of pixels. On the negative side, when a lung is split up into multiple large pieces, such as when the diaphragm enters the CT images at the bottom of the data set, part of the lung will be eliminated because there will be more than two pieces which constitute the lungs on that particular slice. This erroneously excludes part of the lung volume. This also affects the lung volume when it is divided into multiple pieces by vascular tissue. Also, near the top and bottom of the lungs, parts of the lungs will not necessarily exceed the 1% threshold so they will also be incorrectly eliminated from the lung volume. Therefore, this technique, developed by [18], works well for the larger, well defined parts of the lungs but has difficulties near the top and bottom of the lung volumes.

To address these shortcomings region growing approach was introduced which uses a set of seed images. The initial seed image is found by finding the first slide from the top of the CT scan that has at least two objects which exceed the 1% volume criterion, if there are more than two then the two largest ones are retained. Unfortunately, there are infrequent cases where parts of the lungs are cut-off due to high intensity vascular tissue in the lung volume such as illustrated in figures 4.13 and 4.14. In that case the isolated region would be excluded from the lung volume. To address this two more seed images are obtained. The next two slides are chosen as these seed images and the same volume constraint is used to obtain two lungs. This redundancy is an attempt to maximize the chances that at least one seed image is a proper representation of the lung volume for that particular slice. The seed images are further processed using the mathematical morphology operators of dilation and erosion [3], [40], [41]. The purpose behind this approach is to smooth the lung borders and to fill in any indentations. Indentations are often caused by vascular tissue along the lung wall. Any objects along the lung borders have a high chance of being lung nodules so it is very important that the lung volume include these objects. Large vessels near the centre of the CT images also cause large ruts in the lung borders; they also need to be included in the lung segmentation.



Figure 4.12: (a) The lower portion of the lung volume where the diaphragm has entered into the scan for both lungs. (b) shows the results of thresholding (a) producing three pieces that are legitimate parts of the lung volume



Figure 4.13: A CT slice with a significant amount of vascular tissue in the left lung indicated by a circle



Figure 4.14: (a) The result of thresholding figure 4.13 resulting in a piece of the lung being separated by vascular tissue as is illustrated in (b) where the section being cut-off is zoomed in on



Figure 4.15: A CT slice near the top of the lung volume



Figure 4.16: (a) the result of thresholding figure 4.15. (b) The result of retaining only objects which are larger than 1% of the pixel count of the image in (a).

In figures 4.18 (b), it can clearly be seen that the segmentation process to this point has not properly reproduced the border of the lungs from the original image in figure 4.17. This is more obviously seen in figure 4.18 (c) which shows a close-up of the right lung.



Figure 4.17: A CT slice with significant objects near the lung borders



(a)





(c)

Figure 4.18: (a) the results of thresholding the CT slice in figure 4.17. (b) The result of filling, eliminating the background and retaining the two largest objects that satisfy the volume criterion from (a). (c) An enlarged image of the right lung from (b) clearly showing the state of the lung border.

To alleviate this problem morphological closing using a 30mm disc as the structuring element is used. The choice for size of the structuring element arises from the typical maximum agreed upon size for a lung nodule. In extreme cases nodules have exceeded the 30 mm diameter characterization; however, in the vast majority of cases this

structuring element is sufficient. The structuring element can close gaps in the borders up to 30 mm wide. It does a good job of filling in omitted vessels and it is small enough to prevent the two lungs from being joined. The reason the structuring element is circular is to produce curved boundaries after closing since the lung boundary is generally curved.

After morphological closing a complete and accurate segmentation of the lungs is produced. A comparison between the two segmentations and the final results are illustrated by figures 4.20 and 4.21.



Figure 4.19: The results of morphologically closing the objects in figure 4.18 (b) using a disc structuring element with a 30 mm diameter



Figure 4.20: (a) The grey area represents the original segmentation from figure 4.18 (b); the white area represents the new area covered by figure 4.19. (b) is a magnified image of the right lung (left side of image)



Figure 4.21: The results of the (a) pre-morphology and (b) post morphology segmentation of the CT slice in figure 4.17

Having obtained the final seed images the process can proceed onto the rest of the segmentation procedure which deals primarily with region growing. The region growing aspect of this segmentation technique allows it to overcome the weaknesses present in a

pure volumetric approach which are excluding small top and bottom portions of the lungs as well as omitting chunks of the lungs because they are separated by vascular tissue.

Before the actual region growing process begins the rest of the CT slices are also processed in a manner similar to the seed images. All CT images are thresholded using the same CT value as the seed image. After thresholding all non-body objects in contact with the image border are eliminated. All body objects contained within non-body objects are converted to non-body pixels; this is basically filling in the holes within the lungs. At this point the non-seed images are left for the region growing process. It is important to note since there are 3 separate seed images and this process is done in parallel for each seed image then each seed image is a normal slice in 2 of the other processes. This means that the region growing approach actually applies to the seed images due to the parallel and independent nature of the region growing technique.

Having processed all of the images each CT slice consists of multiple non-body pixel objects. For connected components in 3D a 26 point connectivity scheme is used. This means that for a given pixel any other pixel contained within a 3×3 cube centered at the given pixel is connected. The premise behind the region growing approach is that if non-body pixels are connected to the lung volume on an adjacent slice then they are part of the lung volume. This means the non-body object does not fall under volume constraints and there is no limit to the number of objects in a CT slice that can be included in the lung volume. The seed image is the master lung volume that the rest of the volume is grown from. The lungs are grown both up and down. Another attribute of the lung volume analysis also identifies objects in the target slice by which object they are connected to in the seed slice. Morphological closing is applied independently on these groupings to attempt to join objects that are separate but in fact belong to the same object and also to keep objects separate which do not belong together. This particular approach basically refers to keeping the left and right lungs separate as well as rejoining any parts of the same lung that might have been faintly separated during the thresholding process as in figure 4.14. When the adjacent slice to the seed image is done with the morphological closing then it is complete and acts as the seed image for the slice adjacent to it that has not been processed. Once a CT slice with no more attached non-body objects is reached then the process stops. This process is done in both top and bottom

directions starting from the seed image. Once all three of the lung volumes are completed in parallel then the results of each are logically ORed together to produce the final lung volume. The entire lung segmentation process is summarized in figure 4.22. The stages which involve processes which have been developed or created in this system are highlighted.

The system's overall lung segmentation process improves on past processes by including regions in the top and bottom of the lungs which are eliminated in volumetric based approaches. The system is also able to reattach parts of the same lung that are separated by vascular tissue due to the region growing approach. In addition, by using connectedness as an attribute the lungs can be closed separately. This prevents morphological closing from joining the two lungs if they are close together into one large lung volume. Now with the lung volume properly extracted nodule detection stages of the system can be examined.

50





Chapter 5 Lung Nodule Detection

The main focus of this thesis lies in the lung nodule detection portion of the process. In the complete system overview from figure 4.1 the lung nodule detection process encompasses both the second and third stages of ROI Detection and FP reduction. The first part of the process involves acquiring the actual nodule candidates from the lung volume; the second portion involves reducing the number of candidates by eliminating as many FPs as possible.

5.1 Lung ROI Extraction

The approach used for ROI extraction is based on MLT [8], [42-44]. The main reason for using this technique is that it is both easy to control and at the same time has proven to be effective [19], [45]. Initially, a technique based upon the LDM algorithm from [23] was being developed because of the reported potential for performing well in segmenting connected but unrelated objects.

The MLT technique obtains thresholding levels using the CDF of the lung volume. This allows us to adaptively choose our thresholding levels in a way that is relevant to the density values represented in the lung volume. This non-linear approach is an innovation over the technique used by [19] where uniformly spaced intervals were used at preset threshold values. By using adaptive thresholding our technique is more robust to cases where the density values in the lung volumes are irregular and the density values of interest range over different intervals. An example of a CT scan with an irregular density distribution is provided in figure 4.2.

To select the thresholding values two ranges of interest are identified, the nodule range and the vessel range. The nodule range refers to density values that could likely make up nodules. The vessel range identifies very high intensity density values that can belong to well defined blood vessels or nodules alike. This definition implies that the vessel range is a proper subset of the nodule range. From this point on the term density refers to the *grey level* value not the *HU* metric. The *HU* unit provides a primarily negative value range in the lung volume where the *grey level* provides only positive values.

$$Grey \, Level = HU + 1024 \tag{1}$$

The *HU* values range between -1024 and 3072 for a given CT scan where the *grey level* equivalents range between 0 and 4096. Based on the presented range it is easy to infer that the density values in our CT scans are represented by 12-bit values. A number of CDFs of the lung volume from a CT scan can be seen in figure 5.1.



Figure 5.1: Three CDFs of three separate lung volumes



Figure 5.2: Three CT slices coming from the three separate lung volumes represented by the CDFs in figure 5.1 aligned from left to right

In terms of the nodule density range the goal was to look at the range of intensity values where nodules would be likely found. The approach by [22] uses a single threshold at the 80% value from the CDF and gives at least a good starting point within which to experiment for the bottom of the nodule range. Here the grey level values corresponding to a range of CDF values for the three different lung volumes are looked at.

	60%	65%	70%	75%	80%	85%	90%
CT Study 1	191	202	215	232	256	296	374
CT Study 2	294	311	331	357	393	445	528
CT Study 3	335	354	377	407	448	504	586

Table 5.1: The grey level values at different CDF percentage values for the three lungvolumes from figure 5.1

The first thing observed from table 5.1 is that the grey level values can be very different for the same CDF values between lung volumes. These results support the use of the CDF because the CDF is context sensitive to the variations in the intensity values of the lung volumes where absolute grey level threshold values are not. By using the CDF the thresholding is performed relative to the density profile of the particular lung volume [46].



Figure 5.3: the first CT slide from figure 5.2 thresholded at each CDF value from 60% through to 90% from table 5.1 from left to right, top to bottom



Figure 5.4: the second CT slide from figure 5.2 thresholded at each CDF value from 60% through to 90% from table 5.1 from left to right, top to bottom



Figure 5.5: the third CT slide from figure 5.2 thresholded at each CDF value from 60% through to 90% from table 5.1 from left to right, top to bottom

The goal of obtaining the bottom threshold for the nodule range was to have a CT value below which it is unlikely that a nodule would exist. Based on the coverage of the different CT levels from figures 5.3 to 5.5 the threshold was initially chosen to be 65%. This was later chosen to be the bottom of the nodule range. In some cases where the density profile is very high the 65% CDF level would still be too high to detect faint nodules. To address this issue an absolute grey level floor value for thresholding was found and tested against. If the 65% CDF value was above 300 then the lowest thresholding value was set to 300. The levels between the 65% CDF value and the 300 grey level value were uniformly spaced at 20 and were used to only detect nodules. This step size was based upon observed steps in the nodule range during testing and is small enough to provide good segmentation results.

In defining the bottom of the vessel range the main objective is to include only well defined, strongly formed vessels. To satisfy this condition a very conservative approach was taken and a CDF value of 92.5% was chosen for the bottom, this exceeds the highest threshold values presented in figures 5.3 to 5.5. From figures 5.3 to 5.5 it can be seen that using a 90% threshold value already leaves only very high intensity objects, by choosing a higher threshold it ensures a very rigorous standard for what objects can be considered vessels in terms of their density. A high threshold value attempts to prevent separate connected objects from being classified as vessels since connections between objects tend to be of a lower density value than solid individual objects.

The initial threshold value is common for both ranges and is the highest level used in the thresholding process. This level is the starting point for the thresholding process and is higher than both the bottom values of the nodule and vessel ranges. The initial threshold value is chosen as the 97.5% CDF value. Even at 92.5% only very high intensity objects will exceed the threshold and objects are very well separated. The CDF value of 97.5% is a very safe initial threshold in that with very high certainty objects at this threshold will be separate unless they are actually strongly connected, or in fact the same object.

In figure 5.6 it can be seen what the objects look like that exceed the bottom threshold value of 65%. This image clearly illustrates how far the bottom nodule range value of 65% goes towards trying to detect low intensity structures as emphasized in
figure 5.7. Table 5.2 shows that all of the studies listed are close to having the 65% CDF value near 300 or below. In these cases using 300 as a bottom threshold does not significantly change the CDF based approach.

	65%	92.5%	97.5%
CT Study 1	202	444	668
CT Study 2	311	586	768
CT Study 3	354	638	804



Table 5.2: The grey level values at the threshold defining CDF values from figure 5.1

Figure 5.6: The thresholding of the CT slice at the top left corner using CDF values at 65%, 92.5% and 97.5%. The thresholded images retain their original pixel densities where the threshold value is exceeded



Figure 5.7: A comparison of the 65% threshold of the slice from figure 5.6 and the resulting image when the mask is applied to the original CT slice. Of note is how many dark pixels exceed the threshold.

Now that the ranges are defined a set of thresholds is defined. Given the relatively, small CDF difference within the vessel range of 92.5% and 97.5%, 10 evenly spaced threshold levels are used in this range inclusive of range end points. The range between, 65% and 92.5% contains 20 evenly spaced threshold levels including the 65% value but not the 92.5% value which is contained in the previous range. This increase in the number of levels is determined by the more important role that this range plays in the lung volume since it encompasses 27.5% of the CDF. The number of levels and the use of the CDF were determined through experimentation. The various stages and performances using different thresholding approaches are detailed in the results section chapter 6.

One novel component to the MLT approach that has been added was using a density range to obtain one set of ROIs. The premise behind this is that if a nodule is attached to a high intensity object that is denser than itself it will not be separated from that object with traditional thresholding. In this case the nodule range is used, which is defined to be between the 92.5% and 65% CDF value and does not include the absolute bottom mentioned earlier. All pixels that fall within the nodule region are left on while all others are left off. This separates medium intensity objects from high intensity objects. In this implementation it only leaves on pixels present in the nodule range. The implementation of the approach is relatively simple but in its current form it is an

exploratory look at a concept that could be expanded upon and potentially used to produce an entirely new ROI extraction process. Instead of thresholds a range that is moved in steps from the top to the bottom of the density profile of the data set could be used and it would be able to separate objects with density differences. This offers a new and interesting approach to ROI extraction that is worth exploring in future research.

Returning the focus on the implemented ROI extraction process, each level uses an 8 connectivity 2D scheme to identify objects on each slice. Each threshold produces ROI objects. ROI objects in the vessel range are considered for both nodule and vessel candidates. Objects detected in or below the nodule range but not in the vessel range are considered only as potential nodule candidates. This produces thousands to tens of thousands of objects at each level. The MLT approach will process hundreds of thousands of objects for each CT scan due to because at least 24 thresholding levels will exist.

A comparison between our approach and a technique developed by [23] which uses local density peaks to identify ROIs was carried out during the development process. Performance, time and ROI extraction are compared in the results section, chapter 6.

5.2 FP reduction

In order to reduce the number of ROIs a FP reduction system needs to be implemented. The very first step taken to reduce the number of FPs is size filtering. The datasets that are targeted contain 5 mm thick CT slices. In addition, this system initially targeted nodules whose areas are greater than or equal to a circle of 4 mm in diameter (ie. 12.6mm²). In addition, since wall nodules are inherently more suspicious and can be quite small their minimum size needed to be larger than a circle with a diameter of 3 mm (7.1mm²). Experimentation led to making the area range variable as opposed to fixed. The area range now depends upon the X, Y resolution of the CT scan. For scans with an X, Y resolution above 0.65 mm the minimum areas remain at the 3 mm and 4 mm circles for the wall and interior nodules respectively. For scans with a higher resolution, a pixel area of 0.5 is used when calculating the area. The main reasoning behind this was that visually, at a resolution of 0.7 mm, 3 and 4 mm circles represented objects that were

clearly identifiable when looking at the picture. With higher resolutions, smaller objects will appear bigger than they would in a lower resolution scan. Since the 3 and 4 mm circle models worked well for resolutions around 0.7 mm the approximate area of 0.5 mm² was used from these CT images to calculate the minimum areas for higher resolution CT scans. The effectiveness along with the implementation of this approach is further examined in the results section, chapter 6. Size constraints are a very important consideration in the FP reduction process as they are very effective and usually represent the first stage of FP reduction. Using size alone brings down the number of ROIs at each threshold level from hundreds of thousands to a few thousand or hundred. This not only also saves a lot of computational power by eliminating so many candidates early on in the FP reduction process.

The FP reduction system uses two separate approaches to reduce FPs for nodule⁶ candidates with and without contact to the lung walls, the effectiveness of a dual⁶ approach like this has shown to produce good nodule detection results [17], [19].

It is generally more difficult to segment some nodules which are in contact with the lung wall, in particular when they are growing along the lung wall. In these cases the ROI extraction system can usually identify a portion of the lung nodule because of morphological closing. Also, objects in contact with the lung walls are generally more suspicious than interior objects so the system is more sensitive by considering a wider range of wall connected objects through the more permissive minimum area criterion. As an important addition at this point there is one significant problem that arises from the lung segmentation that affects the FP reduction process. When looking at many of the segmented images, a haloing affect around the lung border can be observed. These pixels interfere with feature extraction for the FP reduction phase. This haloing effect can be seen around the lung borders distinctly in figure 5.5. To remove this border a simple erosion of the lung segmentation area is performed using a disc structuring element with radius 1. This erosion is performed prior to any features being calculated for the ROIs, including the area.

it all the states



1. M. P. M. M. M.

和同時間。特許對地位

Figure 5.8: A disc structuring element of radius 1

It can be clearly seen in figure 5.9 that eroding the lung volume helps to significantly reduce the lung border effect in particular at higher threshold values since less of the border is included in the threshold. At this stage the effect is reduced enough to allow for real border objects to be identified. The remaining lung border artifacts are dealt with later on in the FP reduction process or are eliminated during the area filtering procedure. At this stage of the process nodule type classification is explored.



Figure 5.9: The effect of eroding the lung volume on the thresholded images at CDF values of 65% and 80% from figure 5.5

To classify nodules as interior or wall nodules the *perimeter* feature is calculated. The feature is calculated by dilating the object using a disc element of radius 2. By looking at figure 5.10 the disc structuring element is exactly what a single pixel at the centre would look like if it was dilated using figure 5.8 twice in a row.



Figure 5.10: A disc structuring element of radius 2

After dilation, those pixels which are new to the object would be considered perimeter pixels. If any of those pixels lay outside of the lung volume then those are considered to be external pixels.

$$perimeter = P = \frac{p_e}{p_b}$$
(5.1)

The *perimeter*, P, is basically the percentage of surrounding pixels that are outside of the lung volume where p_e are the number of perimeter pixels external to the lungs and p_b are the number of external pixels which are internal to the lungs. It is important to note that even though the surrounding pixels are not all in direct contact with the object, given the structuring element in figure 5.10, they are still at most separated by one pixel. Visually, a spacing of one pixel would usually be interpreted as connected since the object would be very close to the lung border. Therefore, in the worst case some objects would have a *perimeter* value greater than 0 if they are extremely close to the lung wall which could be practically interpreted as being connected. Given these observations the structuring element in figure 5.10 is suitable to use for the *perimeter* feature.

The main use of the *perimeter* value is to check whether or not the object is connected to the lung walls. If the object has a perimeter value greater than 0 it is

considered a wall nodule candidate, if it is equal to 0 then it is an interior nodule candidate. The actual *perimeter* metric is unique to this detection system, as other systems have distinguished between wall and interior nodules before [17], [9] without actually quantifying the strength of the connection as has been done through the *perimeter* feature. This measurement is used further along in the FP process for more than just separating nodules into wall and interior nodules. It also classifies wall nodules as lightly or strongly connected.

The next step in FP reduction proceeds onto identifying nodules and vessels from the wall and interior nodule candidates. Nodules range in shape from spherical to ellipsoidal in terms of 3D shape. In terms of cross-sectional shapes this amounts to nodules being circular to elliptical. The shape of an ellipse is determined by its major to minor axis ratio. A circle is the case where an ellipse has a major to minor axis ratio of one. The shape of an ellipse and a circle also tend to be very compact. One way to evaluate compactness is to look at the area of an object and compare it to the area of its bounding box [17], [23].

From figure 5.11 the bounding box of an object can easily be visualized. However, this bounding box idea does not traditionally work well for angled objects since the bounding box has been predicated on the X and Y dimensions of the object.



Figure 5.11: A couple of ellipse with their accompanying bounding boxes

The difference between a fitted tight bounding box and the traditional X,Y bounding box can be seen in figure 5.12. With the X, Y bounding box both the sides of the rectangle and compactness are dependent on the orientation of the object. With a tight bounding box an object could be at any orientation and the rectangle dimensions and compactness would be the same.



Figure 5.12: An example of what an X, Y based bounding box looks like on the left and a proper, tight bounding box on the right

In an attempt to obtain a tight bounding box the orientation of the object is obtained. The orientation is found as the angle between the x-axis and the ellipse with the same second moments as the object [46], [47].

$$m_{ab} = \iint x^a y^b dx dy \tag{5.2}$$

In (5.2) a and b are constants which determine the order of the moment represented by m_{ab} . Also, x and y refer to spatial coordinates on the x and y axes on a 2D surface. The order of the moment in (5.2) is equal to a + b. The orientation of the object is determined by using (5.3) which defines the orientation of an ellipse based on its second order moments.

$$\tan 2\theta = \frac{2m_{11}}{m_{02} - m_{20}} \tag{5.3}$$

In (5.3) θ is the angle that represents the orientation of the object with second moments m_{11} , m_{02} and m_{20} . Once the orientation is obtained using (5.3) it is then rotated [47] with bi-cubic interpolation [48] applied in order to set its new orientation at 0°. Once the objects orientation is at 0° the bounding box can be obtained. This technique basically amounts to rotating objects such that the trivial X, Y bounding box represents the tight bounding box. Obtaining a tight bounding box for each object is unique to this technique.

Two features are computed using the tight bounding box, the *compactness* [23] and the *axis ratio*.

axis ratio =
$$R_{ar} = \frac{max(L_x, L_y)}{min(L_x, L_y)}$$
 (5.4)

$$compactness = C = \frac{A_o}{A_{bb}}$$
(5.5)

Both of these features relate to shape of the object. In (5.4) L_x refers to the length along the x-axis of the bounding box where L_y refers to the length along the y-axis. In (5.5) A_o refers to the area of an object and A_{bb} refers to the area of the objects bounding box. The *axis ratio*, R_{ar} , looks at the ratio of the long side of the bounding box to that of the short side to get an idea of the 2D symmetry of the object. A long *axis ratio* is usually associated with objects such as horizontal vessels in the lungs. *Compactness*, *C*, is a good identifier of nodules, ellipses or other generally compact shapes. Since nodules tend to appear as circular or elliptical objects on CT slices then a high compactness is another discriminating feature for nodules. Looking at the areas of circles, ellipses and rectangles [47] expected compactness values can be attained.

Area of Circle =
$$A_c = \pi r^2$$
 (5.6)

Area of
$$Ellipse = A_e = \pi ab$$
 (5.7)

Area of Rectangle =
$$A_r = l \times h$$
 (5.8)

Compactness of a circle =
$$C_c = \frac{A_c}{A_r}$$

 $C_c = \frac{\pi r^2}{l \times h}$, where $l = h = 2r$
 $C_c = \frac{\pi r^2}{(2r) \times (2r)} = \frac{\pi r^2}{4r^2} = \frac{\pi}{4} \approx 0.79$

nter skize see tek

Compactness of an ellipse =
$$C_e = \frac{A_e}{A_r}$$

 $C_e = \frac{\pi ab}{l \times h}$, where $l = 2a$, $h = 2b$
 $C_e = \frac{\pi ab}{(2a) \times (2b)} = \frac{\pi ab}{4ab} = \frac{\pi}{4} \approx 0.79$

In the above equations r refers to the radius of a circle, a and b refer to the major and minor axes of an ellipse and l and h refer to the length and height of a rectangle. From the above equations, both ellipses and circles have theoretically the same compactness of 0.79. In practice, since these are represented using discreet matrices and there is no guarantee they are perfectly well formed, due to a nodule's actual shape along with image noise, its *compactness* will vary. The *compactness* and *axis ratio* chosen for the system were based on experimental results and the specific performance of targeted border cases.

Given that lung nodule detection is image based this system used a nodule that was visually judged to be a good border case in terms of *compactness* and *axis ratio*, as identified in figures 5.13 - 5.15. Border cases are important in setting thresholds for separating nodules from non-nodules and in classification problems in general [49].



Figure 5.13: A CT slice with an example of a nodule, circled in black, that was used to shape the *compactness* and *axis ratio* nodule thresholds for the system



Figure 5.14: A threshold of figure 5.13 at the 72.5% CDF level



Figure 5.15: The nodule circled in figure 5.13 and thresholded in figure 5.14

This nodule has a slight L shaped feature which will reduce its *compactness*, showing how an elliptical object can be slightly skewed in a noisy, discreet environment. It also is relatively long and so falls within the extreme case of what an acceptable nodule's *axis ratio* would be. The traditional X, Y bounding box provides a *compactness* of 0.5 and an *axis ratio* of 1.44. Since the nodule is on an angle the compactness is obviously lower then a tight bounding box would represent. Likewise, obtaining the axis ratio is also

inaccurate as 1.44 is too low. After obtaining the tight bounding box the new *compactness* is 0.65 and the *axis ratio* is found to be 1.86. These numbers are more intuitively inline with the visual interpretation of the object in figure 5.14. To allow for some flexibility with these thresholds the system used slightly less strict values than the values calculated for this nodule. The *compactness* threshold was 0.6 and the *axis ratio* threshold was chosen as 2.0. The nodule in figure 5.14 is approximately equivalent to a circle with a diameter of 6 mm. It is important to note that these values are thresholds along with exhibiting a *perimeter* value below 0.05 and an *area* minimum were accepted. The low *perimeter* threshold allows nodules which are weakly connected to still be considered interior nodules. The *area* threshold was examined earlier in the chapter as the minimum size of a nodule that could be detected.

Nodules in contact with the lung wall are handled differently. Objects that exceed the minimum area criterion, in contact with the lung wall, are considered very likely lung. nodule candidates. Wall nodules have slightly more lenient compactness and axis ratio requirements of 0.5 and 2.5 compared to that of an interior nodule candidate. Wall nodules that are attached weakly to the lung wall often can be more elongated and less compact than interior nodules due to the connection, so that is the reasoning behind using a lower compactness value and a higher axis ratio. Weakly connected wall nodules are considered to have a perimeter value of 0.25 or lower, meaning at most 25% of their perimeter pixels are connected to the lung wall. Strongly connected nodules are considered to have a *perimeter* value between 0.25 and 0.65. Strongly connected nodules do not have any compactness or axis ratio restrictions as they can take on irregular. shapes and curvatures that are heavily influenced by the lung wall contour due to their strong connections. The upper limit of 0.65 is never practically surpassed as objects with higher perimeter values would not be segmented since they do not intrude into the lung volume significantly. Even a semi-circle would have a perimeter ratio of approximately 0.61 and that is with the orientation of the curved side touching the lung wall.

It is important to note that all wall nodules have to be considered *thick* as well. This addresses haloing artifacts that remain after lung volume segmentation. Before FP reduction morphology was used to reduce the lung volume to reduce the haloing artifacts

as seen in figure 5.9. Even with this reduction some haloing artifacts remain, and in a further step to eliminate them all wall nodules must be able to be morphologically eroded using the structuring element in figure 5.10 without being eliminated. This process attempts to get rid of any remaining thin artifacts along the lung wall and will not consider them as wall nodules.

Lastly, at this stage of the process the system attempts to identify vessels and vessel groupings. As mentioned earlier, vessels can only be identified in the vessel thresholding range. Objects in this density range must be larger than a 5 mm diameter circle, have a *compactness* of over 0.4, an *axis ratio* of over 2.5 and a *perimeter* value of under 0.05. Vessels are generally very compact so using a minimum *compactness* value of 0.4 is lenient and allows for V-shaped, branching or curved vessels which are less compact than straight, horizontally oriented vessels. A minimum *axis ratio* of 2.5 ensures the vessels are elongated, which is consistent with horizontal vessels. A *perimeter* maximum of 0.05 ensures vessels are not significantly connected to the lung walls, which prevents irregularly shaped wall nodules from being considered vessels.

To address the issue of large groupings of vessels, any objects in the vessel density range which exceed the size of a 20 mm diameter circle and have a *perimeter* value below 0.2 are considered to be large groupings of vessels. Also the *perimeter* value needs to be greater than 0 ensuring contact with the lung wall. The size threshold ensures only large groupings that are unlikely to be nodules are considered. The *perimeter* value is consistent with the idea that vessel grouping branches will skew the *perimeter* towards a lower value since the branches would dominate the perimeter rather than the connection to the lung wall. Also, a maximum *compactness* value of 0.55 is used as vessel groupings are expected to have branches that will tend to lower the *compactness* measurement.

To prevent nodules from getting too big an upper area limit is placed upon them. No interior nodule can be larger than a 15 mm circle. From our data sets only nodules connected to the lung wall ever approached or surpassed this size limit. This prevents large sections or entire slices of the lungs from being considered lung nodules at lower thresholds. A summary of these constraints is presented in table 5.3.

Grouping	A727) Altamater:07 Caputyulant etrigta)	Ebmpartness	avels Rátho	Perfman
Interior Nodule	variable (min) 15 mm (max)	0.6 (min)	2.0 (max)	0.05 (max)
Weakly Connected Wall Nodule	variable	0.5 (min)	2.5 (max)	>0 0.25 (max)
Strongly Connected Wall Nodule	5 mm	-	, ,	> 0.25 0.65 (max)
Vessel	5 mm	0.4 (min)	2.5 (min)	0.05 (max)
Vessel Grouping 20 mm		0.55 (max)	-	>0 0.2 (max)

Table 5.3: A summary of ROI groupings and there corresponding feature thresholds

The next stage of the process looks at all interior nodule candidates that are only one CT slice deep. Any nodule that exists on one CT slice needs good contrast with adjacent slices, otherwise it is most likely connected to a vessel, meaning it itself is most likely a vertical vessel.



Figure 5.16: Three rows of grey level corresponding to nodule vertical contrast analysis.The middle row is grey level values left to right ranging from 1000 to 100 in increments of 100. The range of gray values is between 1200 and 0, corresponding to white and black respectively. The top row is 85% of the middle grey level value. The bottom row is the middle row value minus 100.

For contrast analysis the mean value of adjacent pixels above and below the nodule candidate are compared to the mean value of the nodule's pixels. The contrast value for the adjacent pixel's mean is calculated using the mean of the nodule candidate.

$$Vertical \ Contrast = R_v = \min(N_{mean} \times 0.15, \ 100)$$
(5.9)

In (5.9) N_{mean} refers to the mean pixel value of the nodule. The contrast criterion from (5.9) is visualized in figure 5.15. Given a nodule of the average intensity represented by the middle row of figure 5.15 the *vertical contrast* needs to be at least that of the brighter of the two squares directly above and below the value. It can be seen that this contrast requirement is lenient. The reason for this is to not be overly aggressive in eliminating nodules during this stage. The certitude of the reasoning behind the contrast analysis is that one slice nodules need to be more intense than their vertically adjacent pixels. There is no particular specification of how much contrast is required so a moderate contrast value in (5.9) is used to ensure this elimination rule is not overly aggressive.

The next stage of FP reduction is designed to produce 3-D compact representations of all nodule candidates. The premise behind compaction of the nodule can be seen in figure 5.17.

Figure 5.17: The conceptual idea behind compacting nodule candidates. From left to right the original nodule, nodule with weak connections such as small vessels eliminated, nodule with only significant cross-sectional components retained

First there is a desire to eliminate all weakly 3D connected objects such as small adjacent vessels or small vessels that are feeding blood into the nodule. This is done by eliminating all 2D nodule candidates that are 3D connected to any other nodule candidates who are at least 4 times larger then themselves. Again, this criterion is relatively lenient and strives to ensure that the 2D object being eliminated is relatively insignificant in its 3D context. This first step is the transition between the first and second image in figure 5.17.

The transition from the concept of the middle image to the last image is now performed. This tries to retain the most significant 3D portions of the nodule candidate. This is done by retaining only those 2D components which are within 50% of the area of the largest 2D component for a particular 3D nodule candidate. This processing is attempts to counteract the partial volume effect due to thick slice CT scanning. This stage serves as pre-processing for the rest of the FP reduction process which deals with the nodule in a 3D context. It does this by attempting to retain the most significant portions of each nodule. In addition by eliminating weak connections, it also serves to potentially separate unique 3D objects that were previously connected.

At this stage the vessels identified near the beginning of the FP reduction stage are finally utilized. They are used to process interior nodule candidates. The premise behind this next stage is that if a nodule candidate is engulfed by vessels then it is most likely a vertical vessel. Often a vertical vessel that appears to be a nodule will be covered by a horizontal vessel in an adjacent slice. Also, vertical vessels tend to be represented by a higher grey value when the vessel they branch off of in the same CT slice, so it may be identified as a nodule, but in context it is actually branching vertically from a vessel. It is viewed as a nodule by the system at a higher threshold where the horizontal vessel would not pass. Some contexts for vertical vessels are illustrated through figure 5.18.



Figure 5.18: A CT slice that identifies an isolated vertical vessel on the right and multiple vertical vessels that are branching from horizontal vessels on the left.

This next step looks at any 2D object that overlaps by more than 50% of its pixels with vessel objects in the current and adjacent slices. In figure 5.19 it can be seen how vessels from adjacent slice can cover portions of a potential nodule candidate to identify it as a vertical vessel.



Figure 5.19: The conceptual idea of a potential nodule represented by the circle actually being a vertically vein bridge between two horizontal vessels on adjacent slice. The image shows how projecting vessels onto the current slice covers the majority of the area of the potential nodule.

The threshold for what is considered covered is the majority of the nodule candidate's area, meaning 50% or more. Any 2D object at this point is a significant portion of a 3D object due to the previous stage's compaction and so if the 2D object is eliminated any other slices that are part of that object in a 3D context are also eliminated. This is possible due to the careful and strict 3D trimming that took place in the previous step.

Wall nodules are not eliminated in this process because vessels do not often extend out to the lung wall in any significant manner. Vessels are mainly identified near the centre of the lungs and near the trachea entrance at the lung wall. Therefore, it is highly unlikely that a wall nodule candidate could be covered by vessels. This doubt is what prevents the system from applying this FP elimination process on wall nodules.

The last FP reduction step involves analyzing the depth of any nodule remaining that spans more than one CT slice. This is basically a final attempt to eliminate any vertical vessels that might have slipped through the system to this point. This step uses the 3D equivalent of the *axis ratio*, it is called the *vertical ratio*.

vertical ratio =
$$\frac{(slices - 1) \times slice \ thickness}{max(X \ lengths, Y \ lengths)}$$
(5.8)

The vertical ratio is designed specifically for thick slice CT processing on interior nodules. To account for any potential partial volume artifacts the 3D depth of the object less one slice is used. This is a precautionary measure which keeps in line with the theme of being cautious rather than too aggressive in the FP reduction process. The vertical ratio is basically the depth of the nodule candidate less one slice divided by the longest length of all the tight bounding boxes for each 2D cross section, keeping in mind that this analysis focuses on nodule candidates that are more than one slice thick. One slice objects will have a vertical ratio of zero. Any object which has a vertical ratio greater than 3.0 is eliminated from consideration. This is more permissible than the *axis ratio* which has a cut-off of 2.0. This is again a testament to how careful this system is using 3D features in such a thick slice environment. Wall nodules are not considered for this step because they can have significantly more irregular shapes than interior nodules, in particular when they are large. This is based on observations of CT data sets along with radiologist's diagnosis of nodules in these same data sets.

After this last step wall nodules and interior nodules that have made it to this point are combined into one set and these represent the final nodule set that is output by the system. A summary of the FP reduction process can be seen in figure 5.20.



Figure 5.20: A summary of the FP reduction process

Chapter 6 Experimental Results

This chapter presents all the experimental results that deal with the final output of the system. It also presents various comparisons and results for testing done on different portions of the lung nodule detection system. For experimentation one CT data set that contained 10 synthetic nodules along with 19 CT data sets that contained a total of 239 real nodules were used for testing system performance. Images were acquired using a GE LightSpeed Ultra CT scanner, with a slice thickness of 5 mm in all cases with an effective pixel resolution of 12 bits.

6.1 Lung Segmentation

The lung segmentation process was detailed in chapter 4. The main innovation presented in this technique was that 3D region growing increased the size of true lung volume extracted compared to a standard volumetric based approach [18]. Parts of the lung volume that would have been eliminated due to the volume threshold or separated due to vascular tissue within the lungs are retained.

Study ID	Pixels - Volumetric	Pixels - Region Growing	% pixel difference	# slice s - volumetric	# slices - region growing	# slice difference
2099069	451599	577303	27.84%	21	32	11
2124347	454350	537819	18.37%	25	40	15
2173232	1021103	1132014	10.86%	41	55	14
2045724	1139669	1243561	9.12%	43	53	10
2161056	6 16996	714572	15.81%	29	39	10
1673108	1431887	1582161	10.49%	47	55	8
2110370	567832	636755	12.14%	30	33	3
2174158	724028	838172	15.77%	27	32	5
2118038	1835559	2041767	11.23%	43	52	9
2063289	810363	913602	12.74%	41	48	7

Table 6.1: Statistics comparing volumetric and region growing based lung segmentation

A comparison of the volumetric lung segmentation technique and the region growing technique is presented in table 6.1. A simple and effective way of looking at the performance difference between the two is how many more pixels one includes in the lung volume than the other. It can be seen, from table 6.1, that the region growing approach includes more of the lung volume, usually between 10-15%, than the volume based approach. The number of slides containing a portion of the lung volume is also higher in each case. These results are consistent, since the region growing approach is designed to include smaller lung volumes on a slice.

These results support the idea that region growing is a superior approach to volumetric analysis when it comes to lung volume segmentation. The inherent benefit is much more complete lung segmentation at the top and bottom of the lung volume. There is an additional benefit that any portions of the lung that are cut-off in a CT slice will be retained as long as that volume is 3D connected to the lung volume in an adjacent slice.



Figure 6.1: An example of part of the lung volume near the top of the scan that was retained with the region growing approach but not retained with volumetric based lung segmentation



Figure 6.2: An example of part of the lung volume near the bottom of the scan that was retained with the region growing approach but not retained with volumetric based lung segmentation



Figure 6.3: The tip of the left lung is an example of a portion of the lung volume that is cut-off from the rest of the lung but is retained in the lung volume due to the region growing approach

6.2 ROI Extraction Results

This lung nodule detection system involves a novel approach at MLT that uses the lung volume's density profile to establish the thresholds in an adaptive manner. The technique ensures the thresholds are generated based on the CDF of the particular lung volume.

There is a new approach, a LDM technique proposed by [23] which focuses on local density maximums. It also uses a form of MLT; however, objects are distinguished based on their density profile and connections to other objects. The potential of the approach is that it can separate overlapping structures as long as they were represented by distinct density profiles. A good way of visualizing this is with two mountains, one lower than the other. The peak of the lower mountain will distinguish it from the larger mountain. This has good potential in separating nodules from vessels when they overlap in the CT slice. For a MLT approach this is a challenge and often the attached objects will not be separated, particularly when the nodule is of a lower intensity than the object to which it is attached. Given these facts the MLT that was incorporated in this system was compared to the proposed LDM algorithm.

The first major difference between the techniques was that the LDM approach took approximately eight times longer to run than the adaptive MLT approach used in this system. The running time for the LDM fell in the area of 4 hours where the MLT approach ran in approximately 30 minutes on the same data set. This is a significant difference but it is still acceptable if the results of the LDM are significantly better than the MLT approach used by the system.

The next major failing with the LDM system was that the size threshold was so low that objects would be identified small and fail to properly fill out. On the other hand the MLT technique performs an excellent job of filling out nodule candidates, in particular larger ones. Basically, the MLT approach consistently failed to produce accurate or even acceptable representations of nodules in the CT data that it was tested on, as can be seen in figure 6.4.

e version and the second states and the second s

A SAME A MERICAN AND A SAME AND A



Figure 6.4: A CT slice with ROIs (black spots within high intensity structures) obtained by the LDM algorithm in the top image and the adaptive MLT approach in the bottom image.

The adaptive MLT produces ROIs that are better representations of nodule candidates than the LDM technique. For larger objects it can be seen that the LDM algorithm often does not fill them out, in particular a very large nodule near the top of the

right lung (left side of image) in figure 6.4. The idea behind the LDM approach is promising but at this stage of development a technique that was proven and effective was needed for the ROI portion of the system so the LDM algorithm was not used. Also, by using MLT it made it possible to identify vessels, this would have been extremely difficult, if at all possible, using the LDM algorithm. Most likely, an entirely separate system would have been created alongside the LDM algorithm to detect vessels since the LDM algorithm uses the definition of a nodule in how it segments the image from one threshold to another. The adaptive MLT is simpler in that each threshold is independent of other thresholds; this allows for more flexible processing afterwards.

Based on these observations from testing the two systems the adaptive MLT approach was incorporated. It produces more complete ROI representations within the lung volume and allows for very flexible analysis of ROIs which led to the development of vessel identification and application in FP reduction.

6.3 Lung Nodule Detection Results

The most significant portion of the results section is actual data on how well the system detects nodules. Initially the system was tested using simulated nodules produced by the system described in chapter 3. The simulated nodules were applied onto a single CT series that contained no other nodules.



Table 6.2: The results of applying the lung nodule detection system onto a set of simulated nodules

Testing with simulated nodules the system performed very well. The sensitivity shows that 100% of the nodules were detected. The number of FPs was also very good at

1.46 per slice. The results for the simulated nodules are encouraging, however, these results are somewhat expected since the nodules were well formed and regular. For example, all of the nodules were large enough to satisfy the minimum area criteria of the system. Additionally, the nodules are very well formed since they are synthetic, and therefore are easier to segment.

The data compiled in tables 6.3 through 6.6 consists of sensitivity and FP analysis when the system was applied to 19 real CT data scans. The tables represent the results of the system at four different phases in the development process. The data sets are actually divided into two major sets, those beginning with the letters "ID" and those beginning with the letters "FB". The suffix letter of each set represents the testing phase; "C" represents the first phase and "F" represents the fourth and final phase by using alphabetical order. The "ID" set was the first set that was obtained and used for most of the development until the phase 1 results. The results on the "FB" test set in table 6.3 were basically a blind test at that point having developed the system using the "ID" set only. In addition to blind testing, the "FB" set was used as an indicator at that point to see if the system could help improve the sensitivity of a radiologist's diagnosis.

The experiments and results are obtained as follows. The truth sets were obtained by getting an expert radiologist to examine and mark all of the data sets in each location that he detected a nodule. The "ID" and "FB" test sets were done separately but as entire sets in one session each. When the "FB" set was introduced the system was at the performance level displayed in table 6.3. The radiologist identified 4 additional nodules that were not initially detected. This means originally the set had 56 identified nodules as opposed to the 60 currently listed. This means the system helped to identify 7% more nodules. Three were identified in FB3 and one was found in FB9. The most important thing to note from this result is that the system only performed at 50% sensitivity for the data set at the time of this test as seen in table 6.3 for the "FB" data set. This result implies that even at low sensitivity values the computer aided diagnosis can help improve diagnosis accuracy.

e e

	# of	* #	#		Total	Objects	FPs /	Initial			Vessel	Depth	Wall
Series	nodules	detected	Missed	Sensitivity	Slices	Detected	slice	Interior	Contrast	Trimming	reduced	/ Final	only
FB1C	1	0	1	0.00%	40	62	1.55	Û	0	0	0	0	0
FB2C	0	0	0	N/A	67	95	1.42	0	0	0	0	0	0
FB3C	34	20	14	58.82%	57	124	1.82	18	16	16	16	16	4
FB4C	1	1	0	100.00%	50	70	1.38	1	1	1	1	1	0
FB5C	1	0	1	0.00%	61	112	1.84	0	0	0	0	0	0
FB6C	13	4	9	30.77%	40	116	2.80	2	2	2	2	2	2
FB7C	0	0	0	N/A	37	112	3.03	0	0	0	0	0	0
FB8C	1	1	0	100.00%	53	109	2.04	1	1	1	1	1	0
FB9C	8	4	4	50.00%	59	138	2.27	0	0	0	0	0	4
FB10C	1	0	1	0.00%	35	80	2.29	0	0	0	0	0	0
ID101C	58	50	8	86.21%	65	352	4.65	33	33	33	32	32	18
ID51C	0	0	0	N/A	32	98	3.06	0	0	0	0	0	0
ID52C	103	68	35	66.02%	38	213	3.82	15	15	15	11	11	57
ID53C	13	10	3	76.92%	61	181	2.80	2	2	1	1	1	9
ID54C	1	0	1	0.00%	46	138	3.00	0	0	0	0	0	0
ID55C	0	0	0	N/A	51	81	1.59	0	0	0	0	0	0
ID56C	0	0	0	N/A	56	90	1.61	0	0	0	0	0	0
ID57C	3	2	1	66.67%	55	118	2.11	0	0	0	0	0	2
ID58C	1	0	1	0.00%	44	147	3.34	0	0	0	0	0	0
Set FB	60	30	30	50.00 %	499	1018	1,98	22	20	20	20	20	10
Set ID	179	130	49	72.63%	448	1418	2 .8 8	50	50	49	44	44	86
Total	239	160	79	66.95%	947	2436	2.40	72	70	69	64	64	96

Table 6.3: Results of using the nodule detection system with the MLT thresholds set at the 90% CDF value for the top of the vessel range, 80% for the bottom of the vessel range and 65% for the bottom of the nodule range

	# of	#	#	and a subscription of the	Total	Objects	FPs /	Initial		artellar harilest av	Vessel	Depth	•¥Vall
Series	nodules	detected	Missed	Sensitivity	Slices	Detected	slice	Interior	Contrast	Trimming	reduced	7 Final	only
FB1D	1	0	1	0.00%	40	71	1.78	0	0	0	0	0	0
FB2D	0	0	0	N/A	67	95	1.42	0	0	0	0	0	0
FB3D	34	20	14	58.82%	57	124	1.82	18	16	16	16	16	4
FB4D	1	1	0	100.00%	50	70	1.38	1	1	1	_1	_1	0
FB5D	1	0	1	0.00%	61	112	1.84	0	0	0	0	0	0
FB6D	13	5	8	38.46 %	40	122	2.93	3	3	3	3	3	2
FB7D	0	0	0	N/A	37	109	2.95	0	0	0	0	0	0
FB8D	1	1	0	100.00%	53	109	2.04	1	1	1	1	1	0
FB9D	8	4	4	50.00%	59	140	2.31	0	0	0	0	0	4
FB10D	1	0	1	0.00%	35	80	2.29	0	0	0	0	0	0
ID101D	58	50	8	86.21%	65	415	5.62	33	33	33	32	30	20
ID51D	0	0	0	N/A	32	107	3.34	0	0	0	0	0	0
ID52D	103	76	27	73.79%	38	259	4.82	18	16	16	15	15	61
ID53D	13	10	3	76.92%	61	184	2.85	2	2	1	1	1	9
ID54D	1	0	1	0.00%	46	138	3.00	0	0	0	0	0	0
1D55 D	0	0	0_	N/A	51	83	1.63	0	0	0	0	0	0
ID56D	0	0	0	N/A	56	90	1.61	0	0	0	0	0	0
ID57 D	3	2	1	66.67%	55	120	2.15	Û	0	0	0	Ō	2
ID58D	1	0	1	0.00%	44	164	3.73	0	0	0	0	0	0
Set FB	60	31	29	51.67%	499	1032	2.01	23	21	21	21	21	10
Set ID	179	138	41	77.09%	448	1560	3.17	53	51	50	48	46	92
Total	239	169	70	70.71%	947	2592	2.56	76	72	71	69	67	102

Table 6.4: Results of using the nodule detection system with the MLT thresholds set at the 90% CDF value for the top of the vessel range, 80% for the bottom of the vessel range and 65% for the bottom of the nodule range. In addition if the bottom of the nodule range is higher than 300 then 300 is used as the bottom of the nodule range at steps of 20 between thresholds from the 65% threshold.

	# of 🔬	#	#		Total	Objects	FPs/	Initial			Vessel	Depth	Wall	
Series	nodules	detected	Missed	Sensitivity	Slices	Detected	slice	Interior	Contrast	Trimming	reduced	/ Final	*anly*	
FB1E	1	0	1	0.00%	40	89	2.23	0	0	0	0	0	0	. ,
FB2E	0	0	0	N/A	67	98	1.46	0	0	0	_0	0	0	
FB3E	34	24	10	70.59%	57	133	1.91	14	14	14	14	14	10	
FB4E	1	1	0	100.00%	_ 50 _	91	1.80	1	1	1	1	1	0	
FB5E	1	0	1	0.00%	61	126	2.07	0	0	0	0	D	0	e na L
FB6E	13	7	6	53.85%	40	127	3.00	3	3	3	3	3	4	
FB7E	0	0	0	N/A	37	122	3.30	0	0	00	0	0	0	•
FB8E	1	1	0	100.00%	53	133	2.49	1	1	1	1	1	0	
FB9E	8	6	2	75.00%	59	139	2.25	1	1	1	1	1	5	
FB10E	1	0	1	0.00%	35	106	3.03	0	0	0	0	0	_ 0	
ID101E	58	52	6	89.66%	65	421	5.68	38	37	37	35	34	18	A.
ID51E	0	0	0	N/A	32	113	3.53	0	0	0	0	0	Ō	1
ID52E	103	78	25	75.73%	38	263	4.87	21	19	19		18	60	i .
ID53E	13	10	3	76.92%	61	197	3.07	2	2	1	1	1	9	1
ID54E	1	0	1	0.00%	46	150	3.26	0	_0	0	0	0	0	
ID55E	0	0	0	N/A	51	97	1.90	0	0	0	0	0	0	1
ID56E	0	0	0	N/A	56	116	2.07	0	0	0	0	_0	0	
ID57E	3	2	1	66.67 %	55	120	2.15	0	0	0	0	0	2	
ID58E	1	0	1	0.00%	44	180	4.09	0	0	0	0	0	0	
Set FB	60	39	21	65.00 %	499	1164	2.25	20	20	20	20	20	19	
Set ID	179	142	37	79.33%	448	1657	3.38	61	58	57	54	53	89] .
Total	239	181	_ 58	75.73%	947	2821	2.79	81	78	77	74	73	108	<u>}</u> .

Table 6.5: The same system configuration as in table 6.4 except that now the top of the vessel range is now the 97.5% CDF value and the bottom of the vessel range is the 92.5% CDF value.

.

1

بيا أستاد فرد

	# of	#	#		Total	Objects	FPs/	Initial		an kanadara karangara sar	Vessel	Depth	VVall.
Series	nodules	detected	Missed	Sensitivity	Slices	Detected	slice	Interior	Contrast	Trimming	reduced	7 Final	only
FB1F	1	0	1	0.00%	40	89	2.23	0	0	0	0_	0	0_
FB2F	0	0	0	N/A	67	98	1.46	0	0	0	0	0	0
FB3F	34	25	9	73.53%	57	138	1.98	19	19	18	18	18	7
FB4F	1	1	0	100.00%	50	91	1.80	0	0	0	0	0	0
FB5F	1	0	1	0.00%	61	126	2.07	0	0	0	D	0	0
FB6F	13_	9	4	69.23%	40	142	3.33	5	5	5	5	5	4
FB7F	0	0	0	N/A	37	152	4.11	0	0	0	0	0	
FB8F	1	1	0	100.00%	53	165	3.09	1	1	1	1	1	0
FB9F	8	6	2	75.00%	59	139	2.25	11	1	1	1	1	5
FB10F	1	0	1	0.00%	35	132	3.77	0	Û	Û	0	0	0
ID101F	58	52	6	89.66%	65	421	5.68	38	37	37	35	34	18
ID51F	0	0	0	N/A	32	130	4.06	0	0	0	0	0	0
ID52F	103	85	18	82.52%	38	334	6.55	45	42	42	40	40	45
ID53F	13	10	3	76.92%	61	205	3.20	4	4	3	2	2	8
ID54F	1	0	1	0.00%	46	173	3.76	0	0	Ō	0	0	0
ID55F	0	0	0	N/A	51	94	1.84	D	0	0	0	0	0
ID56F	0	0	0	N/A	56	116	2.07	0	0	0	0	0	0
ID57F	3	2	1	66.67%	55	122	2.18	0	0	0	0	0	2
ID58F	1	0	1	0.00%	44	215	4.89	0	0	0	0		0
Set FB	60	42	18	70.00%	499	1272	2.46	26	26	25	25	25	16
Set ID	179	149	30	83.24%	448	1810	3.71	87	83	82	77	76	73
Total	239	191	48	79.92%	947	3082	3.05	113	109	107	102	101	89

 Table 6.6: The same system configuration as in table 6.5 except now the minimum nodule size is dependent on the cross sectional resolution of the CT slices to allow for smaller nodules being detected on slices that have a higher resolution.

.

. . .

The layout of the information in tables 6.3 through 6.6 is identical. The first column identifies the particular CT data set. At the bottom each separate set is identified as well as a row representing all the sets together. Columns two through four provide nodule information about the CT set and the performance of the nodule detection system. The fifth column is highlighted, and identifies the sensitivity performance for the data sets identified in the first column. The sixth column relates how many CT slices are contained in the data identified in the first column. The seventh column shows how many objects were detected by the nodule detection system. The eighth column reports the FPs per slice for the particular data set. This number equals the number of objects detected less nodules detected, all divided by the number of slices in column six. The last six columns identify how many true nodules are identified as nodule candidates at different stages of the FP reduction system described in chapter 5. The "Initial Interior" candidates relate to nodules that satisfy the minimum area, *compactness* and major to minor axis ratio criteria directly after ROI identification. This refers to nodules that basically pass the first stage of FP reduction. The last column, "Wall only", refers to nodules that were detected because they were identified as wall nodules and were not picked up as interior nodules. The other five columns refer to progressive stages of processing interior nodules in the FP reduction process. The "Contrast" lists how many nodules still remained after vertical contrast analysis for single slice nodules in addition to further area analysis. The "trimming" column shows how many nodules survived the 3D trimming stage of the FP reduction process. The "Vessel Reduced" column refers to nodules that survive the FP reduction stage that uses vessel objects to check for connectivity and eliminate likely vertical vessels that are still considered nodule candidates. The "Depth / Final" column declares which nodules passed the 3D depth analysis, the final stage in nodule FP reduction. The detected nodules are the sum of the "Wall Only" and the "Depth / Final" columns.

The first phase of testing is presented in table 6.3. The performance of the system can best be summarized by the sensitivity and FP performance. The sensitivity over all data was shown to be 67%. The "ID" set performed at 73% sensitivity with 2.88 FPs per slice. If ID101 was eliminated from consideration, which dominates the statistical results since it contains 103 nodules, the sensitivity was 82% (62/76) with 2.79 FPs per slice.

This performance was promising on the ID set especially when looking at the set without ID101. The main reason for optimism was through an analysis of previous techniques and their results. Sensitivity and FP performance for various nodule detection techniques can be seen in table 6.7.

Detection System	Sensitivity range
	(max[FPs per slice], min[FPs per slice]).
Armato et al.	(72% [4.6], 71% [1.5])
Gurcan et al.	(84% [1.74])
Fiebich et al.	(95.7% [0.3], 30% [6.3])
Zhao et al.	(84.2% [5 per set], fail [fail])
Lee et al.	(72% [1.1])
Kanazawa et al.	(90% [0.3])
Paik et al.	(90% [5.6 per set])
Brown et al.	(78% [15 per set])
Our system	(80% [3.05])

Table 6.7: Sensitivity and FP performance from different nodule detection systems [8], [9], [17], [19], [23], [24], [29], [50], [51] including best and worst reported performance if available

The range of reported sensitivities for different systems varies significantly. The work done by [19] shows an improvement of their system in terms of FP performance not due to using a different data set but because they modified their system. The system does perform consistently at around 70% sensitivity. FP performance tends to be very good for systems that have been evaluated using one testing data set. Quite different results are produced when some systems are tested on different data sets. For example [29] reported excellent results in sensitivity and FP performance. When their system was applied to a second data set it performed at unacceptable levels of 30% sensitivity and 6.3 FPs per slice [17] as seen in table 6.7. A similar experience was encountered when applying the technique designed by [23] to the "ID" data set used for testing our lung nodule detection

system. The system actually failed to produce meaningful results as it could not segment nodules properly using the LDM technique mentioned earlier in this chapter. The FP reduction stage also performed so poorly that no meaningful data was presented. The number of nodule candidates for any CT set exceeded 1000 nodules, of which the vast majority were not formed properly due to the failure of the LDM algorithm on the target data set. Even looking at the results by [50] that showed excellent performance their system only found nodules that were equal to or larger than a 6 mm sphere. Given this consideration, the system would perform poorly on many other data sets, including our own, which include many nodules smaller than this criterion. This shows that systems that show very promising results need to be evaluated carefully, and that high sensitivity and low FPs are no guarantee that the system can maintain that performance on different data sets. The system created by [8] reports much more conservative sensitivity and FP numbers, but subsequent testing [19] showed this level of performance was maintained in terms of sensitivity and improved in the realm of FPs. Also, the data set from [19] was perhaps the most similar to the one analyzed by this system as it consisted of 17 data sets and 187 nodules, where our data set has 19 cases and 239 nodules. This data association makes their performance results a good minimum benchmark for this system. Given all of these results and the preliminary state of this system the performance on the "ID" data set was encouraging both in terms of sensitivity and FP performance.

Subsequent testing on the "FB" data set produced significantly worse results than the "ID" data set with a sensitivity of only 50% but an improvement in FP performance at 1.98 per slice. In addition, an examination of the FP reduction stage to stage results is helpful to see how the FP system performed. In total, 8 nodule candidates were eliminated during the FP reduction stage leading to a decrease from a potential sensitivity of 70% (168/239) to an actual sensitivity of 67% (160/239). This is not a significant decrease in sensitivity and the FP reduction process at the initial stage can contain approximately 40 FPs per slice, which is an unmanageable number for diagnosis. Given this, the FP reduction process is necessary and key to producing an effective system for clinical use. The vessel reduction stage was responsible for eliminating most of the nodules during the FP process. This occurred in some cases when nodules overlapped significantly with large veins. Another problem is that the system does pickup the diaphragm near the bottom of the CT data set, as occurred in case ID52 and this eliminated nodules that were overlapped by the diaphragm. To treat this problem the system attempts to not include the diaphragm near the bottom by ensuring the vessel clusters have a *perimeter* value below 0.2 as indicated in table 5.3. This exploits the fact that the diaphragm tends to have a large perimeter value when it is in contact with the lung wall. However, this does not work when the segmented diaphragm is thresholded when it does not have contact with the lung wall. To alleviate this problem phase 4 of the system, in table 6.6, incorporates a greater than 0 *perimeter* feature to attempt to reduce this effect. Regardless, the diaphragm problem in the vessel reduction stage is not going to significantly improve sensitivity so another modification was made for phase 2 of the system.

Given that the sensitivity was weak at the beginning of the FP reduction system, the ROI extraction system was targeted. One observation was that for ID52 there were many nodules that were visibly dense, but fell below the 65% CDF value of the lung volume because the CT scan as a whole had a very high density profile. A check was put into the system to ensure faint nodules were not missed due to the 65% CDF cutoff. If the 65% CDF value was higher than 300, then the threshold levels were augmented in decreasing increments of 20 from the 65% CDF value to 300. The value of 300 was chosen from observing figure 5.15 and concluding that the intensity value of 300 was a reasonable value for what objectively could be considered to have good contrast with a black background. Steps of intensity 20 ensure good coverage at that level and are adequately small to allow for proper detection of faint objects. This is based on the observation of step size and performance in the original approach within the nodule range of thresholds.

The results of implementing this system change in the ROI procedure can be seen in table 6.4. The overall sensitivity of the system improved to 71% (169/239). There were less vessel eliminated nodules at this point for one main reason. There was more of a skew to detecting wall nodules than before, and since wall nodules are not part of the vessel elimination procedure the number of vessel eliminated nodules dropped from 5 to 2. Also the potential sensitivity performance based on the first phase of the FP reduction process is now at 74% (178/239). This is comparable to the difference observed in phase
1 testing in table 6.3 between potential and actual sensitivity. One major observation is that all but one of the new nodules was found in ID52 so even though sensitivity did increase, performance improved mainly due to one CT data set. This is not to say this was an ineffective or useless modification. There was one other CT study, FB6, which also improved by detecting one additional nodule. Regardless of results the modification was one that made sense when the reason for it was examined, in particular when considering that it is important to account for CT studies that have irregular density profiles.

Phase 3 was developed with more of a focus towards the "FB" data set. On of the major problems with ROI candidates was that some high intensity nodules were not being picked up. Due to this the nodule and vessel ranges were reconsidered and new threshold CDF values were used for the top and bottom of the vessel range. The 92.5% and 97.5% CDF values were used rather than the 80% and 90% values in the previous two results. Likewise, only 4 threshold levels were previously used in the vessel range, producing steps between thresholds of over 50 in some cases. To alleviate this problem the vessel range was chosen to be more compact and the number of threshold levels increased to 10. The higher thresholds also ensure a higher standard for the density and connectivity of objects that are considered vessels. The previous range was too lenient in terms of the density allowance and some diffusion between objects satisfied the vessel density criterion. With the higher thresholds there is less of a chance that these diffusions can be included as vessels. Also, the higher vessel range and increased inter-level resolution ensures that the third phase of development would do a better job of identifying nodules which are visually distinct but exist in a high intensity background or are surrounded by high intensity objects. The performance of the system at the third phase of development can be seen in table 6.5.

The sensitivity results had a decided increase in performance again, now performing at 76% in total. This is a 5% increase over the previous phase, which itself had a 4% increase over the previous phase. Therefore the total sensitivity improvement is consistent with the previous modification on a global sense; however there is a distinct difference. The second phase improved sensitivity by basically improving performance in case ID52 while only improving the "FB" sensitivity from 50% to 52%. The sensitivity for the third phase in the "FB" data set was at 65%, an improvement of 15% and 13%

over the previous results. This is important because the "ID" set already had relatively good sensitivity, even after the first phase of testing but the "FB" set performed very poorly. Also, the "ID" set contains 179 nodules where the "FB" set only contains 60, meaning the "ID" test set dominates global sensitivity performance. Of the 12 new nodules detected, 8 of them were from the "FB" set. The significant improvements made in the FB set in terms of detection sensitivity are a very encouraging observation that strongly supports the new MLT approach towards the vessel range implemented in the third phase of the system.

Looking further into the results presented in table 6.5, the potential sensitivity at the beginning of the FP reduction stage is 79 % (190/239) compared to the actual sensitivity of 76% (181/239). No one particular FP reduction step is responsible for eliminating an exceptional number of nodules so the FP process seems relatively balanced, and has only a small affect on sensitivity.

Given the good sensitivity performance produced by the system after making modifications primarily to the MLT process, the FP reduction stage was now modified for the fourth and last phase of the system, the results of which are shown in table 6.6. One observation, when looking at the results was that some nodules that were well formed did not get detected and it was concluded that the MLT process was not responsible, especially given the modifications made to this point to the thresholding approach. The system was calibrated to accept nodules that were greater than a 4 mm disc in size within the interior and a 3 mm disc along the lung wall. It was found that the reason for this was that the nodules that were not being detected were falling below the minimum size criteria, when other nodules in other CT scans that were smaller than them visually were being detected. Any past literature on the topic of nodule detection always dealt with a lower bound on nodule size in terms of an absolute area as our system has been using to this point. The truth of the matter is that when a human observes the nodules in a CT scan they are not concerned about the geometric area; they actually focus on the pixel area of the object. What happened with the nodule that was not being detected was that the X, Y resolution of the CT scan was very high, making smaller nodules appear visually larger since the pixel area was significantly smaller. Analysis of the system found that many CT scans had X, Y resolutions that were in the range of 0.7

mm. In the case where the nodule was not being detected the X, Y resolution was actually approximately 0.5 mm. This means that smaller nodules visually large but geometrically small. The minimum area criteria were established using the visual feedback of a 0.7 mm CT scan, so to adjust the criteria to be sensitive to the size of the object in terms of pixel a modification was made to the lower area criteria. It can be seen in table 6.8 that a CT scan with an X,Y resolution of 0.5 mm compared to that of one with a resolution of 0.7 mm actually means that objects had to appear twice as big visually to have the same area.

X,Y resolution	Pixel Area	3 mm disc pixel area	4 mm pixel disc area
0.7 mm	0.49 mm^2	15 pixels	26 pixels
0.5 mm	0.25 mm^2	29 pixels	51 pixels

Table 6.8: A comparison of the effect on visual size by the X,Y resolution of CT

scan

Given that the pixel area of a 0.7 mm resolution CT scan was approximately 0.5 mm^2 , and that provided good visual resolution, the pixel area was artificially enhanced to that value when calculating the 3 and 4 mm disc area for high resolution CT scans with an X,Y resolution smaller than 6.5 mm. This stops short of making the pixel area absolute, but it does address the need to be sensitive to the visual representation of the nodule. Given the past systems were sensitive to nodules as small as a 3 mm disc this new criteria lowered the minimum sized nodules that the system could detect. The highest resolution CT data set in the testing data was ID52 with an X, Y resolution of 0.4805 mm. This means by modifying the pixel area the new effective area minimums of potential nodules were actually the equivalent areas of 15 pixel and 26 pixel objects which represent the area of a 3mm and 4mm disc respectively. This means the new minimum area for a wall nodule was 3.46 mm² which is equivalent to a disc with a diameter of 2.1 mm and the new area of for an interior nodule was 6 mm² which is the same as a disc with a diameter of 2.8 mm. This means that the system was now setting the minimum size criteria of what could be considered a nodule based on the X, Y pixel resolution of the CT scan. This is sensible when considering higher resolution CT scans produce larger and better defined

representations of smaller nodules. Our CT scans ranged in resolution between 0.7 mm and 0.5 mm approximately. This modification represents the first nodule detection system that reports using variable size criteria based on the CT scan resolution and additionally represents only the third system that reports being able to detect nodules below 3 mm in diameter [17], [51].

The results of the modification produce a significant improvement in the results, displayed in table 6.6. The system sensitivity increased to 80% (191/239), boosted by a fairly balanced increase from both test sets. It is important to note that this modification only improved the performance of the CT scans which had a high X, Y resolution. From table 6.6 they can be seen to be FB3, FB6-8, FB10, ID51-55 and ID58. This means for CT scans where the modification took effect the old sensitivity of 72% (120/167) improved to 78% (130/167). That represents 10 additional nodules and a sensitivity jump of 6%, which is higher than the total sensitivity jump of 3%. This was a very effective modification to the system that clearly performed well in the cases that were targeted. Now with an overall sensitivity of 80% the system has improved dramatically from the first phase of development which had a sensitivity of 67%. What is even more impressive is that the potential sensitivity of the system when looking at the beginning of the FP reduction process is 85% (202/239), much higher than the previous value of 79% in the third phase of testing.

Another very promising assessment of performance comes from looking only at nodules that satisfy the minimum area criteria of the system, in this case the total number of nodules drops to 211 from 239. This means the sensitivity of the system to nodules that satisfy the minimum area criteria is 91% (191/211). These are excellent numbers in terms of sensitivity and shows excellent potential for detecting nodules.

Using the value of 91% when comparing to any other reported system this provides a very favourable comparison and actually provides a good argument that other systems perform at a lower sensitivity than our system does based on the data sets they have used. For example, a system such as the one proposed by [9] reports sensitivity of 90%, which from table 6.7 represents the pinnacle of nodule detection sensitivity performance. However, it only reports detecting nodules which are larger than a disc with a diameter of 4 mm or greater. Given that only 211 nodules meet the modified size

criteria in phase 4, and that phase 4 had detected 13 more nodules than phase 3, when only the size criteria was altered, it is safe to assume that those thirteen nodules fell below the previous size criteria. That means, of the 239 nodules, it is reasonable to conclude that 41 nodules fall below the 4 mm diameter disc size criteria. This automatically means the system proposed by Kanazawa et al. would miss 41 nodules in our test set purely due to its size constraints. They also claim that they detected 95% of nodules before their FP reduction process which implemented the area cut-off criteria. It can be assumed, in the most extreme case, that all of their nodules were eliminated due to them being too small. This means that 5% of their data set consisted of nodules below 4 mm in diameter. Also, assuming the 5% of their data set that was not picked up by their ROI process was above 4 mm in diameter then 95% of their data consisted of nodules above 4 mm in diameter. If they detected 90% of all nodules where only 95% were detectable due to the size criteria then their system worked at 94.7% sensitivity above 4 mm. Given that our test set consists of at most 198 nodules above 4 mm in diameter their system would have detected 188 nodules if it operated at a sensitivity of 94.7%. This means that at its best, their system could have performed at a total sensitivity of 78.6% (188/239). This is a lower sensitivity than was obtained by our system. It is also important to note that this is a best case assumption for their system. A similar argument can be made for other systems that report higher sensitivity but use a higher minimum area cut-off.

Looking at the results of [51], they used thin section CT which makes detecting small nodules easier. Also, their system had an overall detection sensitivity of 83.5% on a test set of 79 nodules, which is significantly smaller than our test set of 239 nodules. Since our system is operating on thicker slices the detection sensitivity should be lower than on a system with thinner slice CT scans where there is significantly less inter-slice interference. The study by [17] reports a sensitivity of 84% using a data set of 63 nodules, again a significantly smaller data set than our own. If the performance of our system using a subset of 60 nodules from 8 CT data sets in table 6.6 was analyzed the system performs at a sensitivity of 90% (54/60) which is higher than the reported results by [17] but with a comparable data set size in terms of nodules.

化化化化化化化化化化化化化化化化化化化化化化化化化化化化化化化化化化化化

Given the results, our system does perform very well in sensitivity and the argument can and has been made that it has class leading sensitivity based on the results and test set when compared to the test sets and results of other reported techniques.

The major deficiency in our system is the FP performance which is at 3.05 FPs per slice. The FP performance does range between 1.46 and 6.55 FPs per slice so the actual range of FP results on a case by case basis can be competitive when compared to other techniques listed in table 6.7. Likewise, due to the diverse nature of our test set it is difficult to compare the FP performance of other systems given how performance is highly dependent on the actual CT data sets as our own data has shown given the intra-set range of FP results. The challenges of the test set aside, the system's FP performance is acceptable at this stage and there is a focus to improve it with further development.

and the second of the second . and the second a a cara a c a a contrat na se a completa e la servició de la 了,你们们们的你们,你们的你们都是你们,你们都不是你的。"

Chapter 7 Conclusions

Creating this system was a significant challenge and a great opportunity at the same time. This system has the opportunity to lead the way in automatic lung nodule detection in Canada as no other Canadian based research in this field has been found yet. The greatest challenge of creating a lung nodule detection system was the scope of the project. The main objective of this work was to create a fully functional system that was practical and effective.

en la martin servicio principale de la servicio de

One of the largest obstacles to overcome was the sheer volume of work that needed to be done to create a complete system. Everything was built from scratch and getting started required finding a direction that could be followed.

Looking at the state of the art in lung nodule detection research there are many different approaches but it is difficult to judge the merits of the systems without access to them or their data sets. This makes choosing a path difficult because there are no guarantees that the chosen approach is going to be successful. The first phase of this research demonstrated this as a system that had excellent potential, based on the presented results, was implemented but performed unacceptably on our CT data. In retrospect, with a more critical eye and more understanding of lung nodules, the results of the system that was initially examined [23] are not surprising. This was a lesson to look deeper than what is presented from the direct performance data.

At this stage of the process a completely functional lung nodule detection system has been developed that meets the initial design goals and was created in a very short period of time. In addition a lung nodule synthesis tool was created that was used in the testing process and additionally provided insight into nodules that helped to develop the rest of the system. From this point forward future research work will have this system as a basis and a springboard. Each of the systems can be examined individually and with more detail. In addition, the techniques that have been implemented in this system can be further refined and examined. Most importantly, new approaches can be carried in the context of the system with more narrow focus and detail since the creation of all the components does not need to be a concern.

The results of the nodule detection system are very encouraging and show good progress was made during the development of the system. The results rose from a 67% total case sensitivity to 80%. Also the sensitivity of the "FB" test set rose from 50% up to 70%, which was another very encouraging sign. In addition, it has been shown that the system's sensitivity performance is class leading when the context of the CT data is applied to the reported results of other techniques. In particular, the number of nodules and the variety of cases makes our CT data set the most robust set that any system has been reportedly tested on.

In addition to excellent sensitivity performance and a robust data set there are many innovations present in the system that can be built upon and applied to other systems. First, a unique 3D region growing approach is used for lung segmentation that improves over volumetric approaches by including more of the lung volume, in particular near the top and bottom of the lungs. Next there is a unique MLT approach that selects thresholds based on the CDF values of the density profile of the lungs. It is the first ROI acquisition system that tries to detect vessels in addition to nodules and uses them in FP reduction; this is done by uniquely defining a vessel range among the thresholds used in ROI extraction. The system is only the second system that uses thick slice CT scans to find nodules below 3 mm in diameter, the other being the one developed by Gurcan et al. [17]. In addition this is the only system that adaptively determines the minimum area criterion based on the X, Y resolution of the image, taking advantage of CT scans with high cross-sectional resolutions to identify small nodules. One of the most significant innovations involves using a tight bounding box to calculate 2D shape features, where, in past systems the orientation of the nodule would interfere with accurate feature extraction. The system itself is of a modular design, permitted through a rule based approach. This allows for future development where rules and stages can be plugged in and out at different points of the FP reduction process.

Given the new ideas presented, and the excellent sensitivity performance on the data sets, development on the system must be continued. Even with excellent preliminary results, comparing systems that were tested using different data sets still prevents an absolute comparison; it only allows analysis and interpretation that forms performance conclusions as has been done. It can definitely be said that this system is better than what others have proposed but it can be inferred that it appears to have the best sensitivity performance among these systems when the context of the data sets is considered. One step towards producing a semblance of absolute comparison is the development of a public nodule database [28] that will allow direct comparison to some degree between systems. The database is intended to contain 400 nodules within thin slice CT scans. Not only will this help systems to be compared but it will also allow research in this area to flourish as any researcher will have access to a large database of real lung nodules.

7.1 Future Work

There is much opportunity for future development on this system as it currently stands. Developing a lung nodule detection system is a large task with many components. Lung volume segmentation, ROI extraction and FP reduction can all be very intricate projects of their own, but piecing all three together into a complete system while developing all them in parallel leaves the door open for improvements in all three areas.

The automated lung segmentation process is an evolution from previous approaches and performs well in our CT data, but further refinement is possible. The trachea is a significant object in the CT scans and further looking at how to segment it separately and analyze how it enters the lungs can lead to improving how they are segmented in those slices. The heart is a significant object that also intrudes on the segmentation and developing techniques that identify it and exclude it from the lung volume can potentially improve the segmentation process. In addition, identifying the bones around the lungs could help the process significantly as they could be used to identify shadows on adjacent slices that increase the number of FPs.

The ROI process is a crucial portion of the lung nodule detection system. A major focus of this work was finding the right thresholds and regions to use in the system.

Improving the ROI extraction process was shown to increase sensitivity by approximately 10% in chapter 6. Exploring other potential systems is a crucial area of future research, as the ROI extraction process may be the most important stage of a lung nodule detection system. Properly extracting the nodules is crucial to allow the system to analyze the nodules the same way a radiologist would when looking at the picture. It is relatively easy for a person to separate attached nodules from a vessel but not so for an algorithm. Even though the LDM algorithm presented earlier failed in its implementation the idea behind it is a good one. Modification and development of that algorithm can potentially lead to properly extracting nodules which are strongly attached to other vascular tissue, in particular when the intensity of the nodule is lower than the attached object. This scenario is difficult for a pure thresholding approach as the nodule can not be separated from the attached object. One potential thresholding approach that could work for such a system would be using a density range rather than a threshold. All pixels within the range would be turned on and the range can move through the density profile similarly to a threshold. The step between ranges and the size of the range would be crucial development decisions. This system has taken a step in this direction by using the nodule range of the threshold steps in this manner during the ROI extraction process.

The FP reduction stage is crucial in the performance of the system and is dependant on a reliable ROI extraction process. Many new and effective techniques have been implemented at this stage of development but FP reduction is the main area that needs to be improved upon within this system. Commercial systems claim FP performance anywhere in the range of 8 to 30 nodules [52]. This system has significant inter study variance in FP performance and in cases is competitive with other research systems. It is again difficult to absolutely compare, in particular given how performance varies among systems as illustrated in chapter 6. However, given the excellent sensitivity results so far, reducing the number of FPs in the future is the main focus. By making the system rule based it was being setup for such future work. FP reduction stages can be changed independently and easily. One major modification that can be made is applying a LDA stage which has shown to be very promising in other systems [17], [19]. Developing such a system would involve a deep analysis of the feature spaces of lung nodules and is a major undertaking that can potentially pay significant dividends in FP

reduction. Also, many features can be examined further such as the *perimeter* feature that is used to identify interior and wall nodules. Adding new features to the system, testing their effectiveness and analyzing the distribution of these features among nodules and FPs is a time consuming but crucial part of further developing the system.

Another consideration in the future will be applying the system to thin slice CT scans. The system is setup in a general way and can be applied directly to thin section CT scans, however it is designed to be very cautious with 3D features given the inaccuracy in 3D metrics produced my thick slice CT scans. Adapting the system to thin slice CT scans will allow it to take advantage of cleaner and more accurate nodule representations potentially producing entirely new features to analyze in the FP reduction process.

The system has a very bright future and there are many important areas that can be developed. A fully functional system with innovative components has been produced in a very short period of time, and there is great anticipation in exploiting its future potential through more research.

Bibliography

ne de la sectore de la se Sectore de la s

[1] T. Yamamoto, Y. Ukai, M. Kubo, N. Niki, H. Satou and K. Eguchi, "Computer aided diagnosis system with functions to assist comparative reading for lung cancer based on helical CT image", The Proceedings of the International Conference on Image Processing, Vol. 1, pp. 180 – 183, Sept. 2000

[2] D. Lee, "Some computational aspects of low-level computer vision", Proceedings of the IEEE, Vol. 76, Issue 8, pp. 890 – 898, Aug. 1998

[3] R. C. Gonzalez, R. E. Woods, <u>Digital Image Processing</u>, 2nd ed., New Jersey: Prentice Hall, 2002

[4] K. Phoi, "Current status and future potential of computer-aided diagnosis in medical imaging", British Journal of Radiology, Vol. 78, pp. 3-18, Jan. 2005

[5] G. Coppini, S. Diciotti, M. Falchini, N. Villari and G. Valli, "Neural networks for computer-aided diagnosis: detection of lung nodules in chest radiograms", IEEE Transactions on Information Technology in Biomedicine, Vol. 7, Issue 4, pp. 344 – 357, Dec. 2003

[6] M. G. Penedo, A. Mosquero, M. J. Carreira and D. Cabello, "Lung nodule detection in curvature space with multilayer perceptron network", Proceedings of the 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol. 3, pp. 1130 – 1132, Nov. 1996

[7] M. G. Penedo, A. Mosquero, M. J. Carreira and D. Cabello, "Computer-aided diagnosis: a neural-network-based approach to lung nodule detection", IEEE Transactions on Medical Imaging, Vol. 17, Issue 6, pp. 872 – 880, Dec. 1998

[8] S. G. Armato III, M. L. Giger, C. J. Morgan, J. T. Blackburn, K.Doi, and H. MacMahon, "*Computerized Detection of Pulmonary Nodules on CT Scans*", Imaging and Therapeutic Technology, RSNA, pp. 1303-1311, 1999

[9] K. Kanazawa, Y. Kawata, N. Niki, H. Satoh, H. Ohmatsu, and R. Kakinuma, "Computer-aided Diagnosis for Pulmonary Nodules Based on Helical CT Images", Proceedings of the International Conference on Pattern Recognition, Vol. 2, pp. 1683-1685, Aug. 1998

[10] F. Li, S. Sone, H. Abe, H. MacMahon, S. G. Armato III and K. Doi, "Lung Cancers Missed at Low-Dose Helical CT Screening in a General Population: Comparison of Clinical, Histopathological, and Imaging Findings", Radiology, Vol. 225, No. 3, pp. 673 – 683, December 2002

[11] D. Sortini, K. Maravegias and A. Sortini, "*Difficulty of early diagnosis in patients with solitary pulmonary nodule*", Journal of Thoracic and Cardiovascular Surgery, Vol. 129, Issue 5, pp. 1196-1196, May 2005

[12] J. W. Lee, J. M. Goo, H. J. Lee, J. H. Kim, S. Kim and Y. T. Kim, "The potential contribution of a computer-aided detection system for lung nodule detection in multidetector row computed tomography", Investigative Radiology, Vol. 39, pp. 649–655, Nov. 2004

[13] Y. Kawata, N. Niki, H. Ohmatsu and N. Moriyama, "*Example-based assisting approach for pulmonary nodule classification in three-dimensional thoracic computed tomography images*", Academic Radiology, Vol. 10, pp. 1402 – 1415, Dec. 2003

[14] American Cancer Society, <u>Cancer Facts and Figures 2005</u>, Atlanta: American Cancer Society, 2005

[15] <u>National Cancer Institute of Canada: Canadian Cancer Statistics 2004</u>, Toronto, Canada, 2004

[16] M. S. Brown, M. F. McNitt-Grey, J. G. Goldin, R. D. Suh, J. W. Sayre, and D. R. Aberle, "*Patient-Specific Models for Lung Nodule Detection and Surveillance in CT Images*", IEEE Transactions on Medical Imaging, Vol. 20, No. 12, pp. 1242-1250, December, 2001

[17] M. N. Gurcan, B. Sahiner, N. Petrick, H. Chan, E. A. Kazerooni, P. N. Cascade, and L. Hadjiiski, "Lung Nodule Detection on Thoracic Computed Tomography Images: Preliminary Evaluation of a Computer-Aided Diagnosis System", Medical Physics, Vol. 29, No. 11, pp. 2552-2558, November 2002

[18] S. Hu, E. A. Hoffman, and J. M. Reinhardt, "Automatic Lung Segmentation of Accurate Quantitation of Volumetric X-Ray CT Images", IEEE Transactions on Medical Imaging, Vol. 20, No. 6, pp. 490-498, June 2001

[19] S. G. Armato III, M. L. Giger and H. MacMahon, "Automated detection of lung nodules in CT scans: preliminary results", Medical Physics, Vol. 28, pp.1552 – 1561, Aug. 2001

[20] M. L. Giger, N. Ahn, K. Doi, H. MacMahon and C. E. Metz "Computerized detection of pulmonary nodules in digital chest images: use of morphological filters in reducing false-positive detections", Medical Physics, Vol. 17, No. 5, pp. 861 – 865, Oct. 1990

[21] Y. Lee, T. Hara, H. Fujita, S. Itoh and T. Ishigaki, "Automated Detection of Pulmonary Nodules in Helical CT Images Based on an Improved Template-Matching Technique", IEEE Transactions on Medical Imaging, Vol. 20, No. 7, July 2001 [22] D.T. Lin, and C.R. Yan, "Lung Nodules Identification Rules Extraction with Neural Fuzzy Network", Proceedings of the International Conference on Neural Information Processing, Vol. 4, pp. 2049-2053, Nov. 2002

[23] B. Zhao, G. Gamsu, M. S. Ginsberg, L. Jiang, and L. H. Schwartz, "Automatic Detection of Small Lung Nodules on CT Utilizing a Local Density Maximum Algorithm", Journal of Applied Clinical Medical Physics, Vol. 4, No. 3, pp. 248-260, Summer 2003

[24] Y. Lee, T. Hara, H. Fujita, S. Itoh and T. Ishigaki, "*Nodule detection on chest helical CT scans by using a genetic algorithm*", Proceedings of the conference on *Intelligent Information Systems*, pp. 67 – 70, 1997

[25] <u>Fuzzy Logic Toolbox</u>, [online document] March 2005 [cited 2005 June 6] Available at: http://www.mathworks.com/access/helpdesk/help/pdf_doc/fuzzy/fuzzy.pdf

[26] M. Nachtegael and E. E. Kerre, "Fuzzy mathematical morphology: general concepts and decomposition properties", Third International Conference Knowledge-Based Intelligent Information Engineering Systems, pp. 255 – 258, Sept. 1999

[27] A. T. Popev, "Morphological operations on fuzzy sets", Fifth International Conference on Image Processing and its Applications, pp. 837 – 840, July 1995

[28] "Lung Imaging Database Consortium", [online document] [cited June 6 2005] Available at: http://imaging.cancer.gov/programsandresources/InformationSystems/LIDC/print

[29] M. Fiebich, C. Weitholt, B. C. Renger, S. G. Armato, K. R. Hoffman and S. Dietrich, "Automatic detection of pulmonary nodules in low-dose screening thoracic CT examinations", Proc. SPIE 3661, pp.1434 – 1439, 1999

[30] J. P. Ko, H. Rusinek, E. L. Jacobs, J. S. Babb, M. Betke, G. McGuinness and D. P. Naidich, "Small pulmonary nodules: volume measurement at chest CT--phantom study", Radiology, Vol. 228, No. 3, pp. 864 – 870, Sept. 2003

[31] L. Pastor, P. Manzoni, M. Parmentier and B. Kastler, "Lung modeling for nodule analysis in HRCT", 2001 IEEE Nuclear Science Symposium Conference Record, Vol. 3, pp. 1504 – 1508, Nov. 2001
[32] W. J. Kostis, A. P. Reeves, D. F. Yankelevitz and C. I. Henschke, "Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images", IEEE Transactions on Medical Imaging, Vol. 20, Issue 10, pp. 1259 – 1274, Oct. 2003

[33] Y. Kawata, N. Niki, H. Ohmatsu, R. Kakinuma, K. Mori, K. Eguchi, M. Kaneko and N. Moriyama, "Quantitative surface characterization of pulmonary nodules based on thin-section CT images", IEEE Transactions on Nuclear Science, Vol. 45, Issue 4, pp. 2132 – 2138, Aug. 1998

[34] Y. Kawata, N. Niki, H. Ohmatsu, M. Kusumoto, R. Kakinuma, K. Mori, H. Nishiyama, K. Eguchi, M. Kaneko and N. Moriyama, "Internal structure analysis of pulmonary nodules in topological and histogram feature spaces", Proceedings of the International Conference on Image Processing, Vol. 1, pp. 168 – 171, Sept. 2000

[35] Y. Kawata, N. Niki, H. Ohmatsu, M. Kusumoto, R. Kakinuma, K. Mori, K. Eguchi, M. Kaneko and N. Moriyama, "*Curvature based analysis of internal structure of pulmonary nodules using thin-section CT images*", 1998 The Proceedings of the International Conference on Image Processing, Vol. 3, pp. 851 – 855, Oct. 1998

[36] R. S. Saunders Jr., E. Samei and C. Hoeschen, "Impact of resolution and noise characteristics of digital radiographic detectors on the detectability of lung nodules", Medical Physics, Vol. 31, pp. 1603 – 1613, June 2004

[37] J. Burns, L. B. Haramati, K. Whitney and M. N. Zelefsky, "Consistency of reporting basic characteristics of lung nodules and masses on computed tomography", Academic Radiology, Vol. 11, pp. 233 – 237, Feb. 2004

[38] M. S. Smyczynski, H. C. Gifford, T. H. Farncombe, W. P. Segars, B. M. W. Tsui and M. A. King, "Impact of respiratory motion on the detection of solitary pulmonary nodules with SPECT tumor imaging of NeoTect", Vol. 2, pp. 838 – 841, Nov. 2002

[39] I. Middleton and R. I. Damper, "Segmentation of magnetic resonance images using a combination of neural networks and active contour models", Medical Engineering and Physics, Vol. 26, Issue 1, pp. 71 – 86, Jan. 2004

[40] S. S. Wilson, "*Theory of matrix morphology*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.14, Issue 6, pp. 636 – 652, June 1992

[41] J. Yang and X. Li, "Directional morphology and its application in boundary detection", Fifth International Conference on Image Processing and its Applications, pp. 742 – 746, July 1995

[42] J. R. Parker, "Grey level thresholding in badly illuminated images", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 13, Issue 8, pp. 813 – 819, Aug. 1991

[43] J.-C. Yen, F.-J. Chang and S. Chang, "A new criterion for automatic multilevel thresholding", IEEE Transactions on Image Processing, Vol. 4, Issue 3, pp. 370 – 378, March 1995

[44] N. Ramesh, J.-H. Yoo and I. K. Sethi, "*Thresholding based on histogram approximation*", IEE Proceedings on Vision, Image and Signal Processing, Vol. 142, Issue 5, pp. 271 – 279, Oct. 1995

and in the second second of the second of the second second second second second second second second second s If the second [45] S. G. Armato III, M. B. Altman and P. J. La Rivière, "Automated detection of lung nodules in CT scans: Effect of image reconstruction algorithm", Medical Physics, Vol. 30, No. 3, pp. 461 – 472, March 2003

[46] A. Leon-Garcia, <u>Probability and Random Processes for Electrical Engineering</u>, 2nd ed., Massachusetts: Addison-Wesley, 1994

[47] H. Anton, C. Rorres, <u>Elementary Linear Algebra Applications Version</u>, 7th ed., New Jersey: Wiley, 1994

[48] R. Keys, "*Cubic convolution interpolation for digital image processing*", IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 29, Issue 6, pp. 1153 – 1160, Dec. 1981

[49] W.A.H. Mousa, and M.A.U. Khan, "Lung Nodule Classification utilizing Support Vector Machines", Proceedings of the International Conference on Image Processing, Vol. 3, pp. 153-156, June 2002

[50] D. S. Paik, C. F. Beaulieu, G. D. Rubin, B. Acar, R. B. Jeffrey Jr., J. Yee, J. Dey and S. Napel, "Surface normal overlap: a computer-aided detection algorithm with application to colonic polyps and lung nodules in helical CT", IEEE Transaction on Medical Imaging, Vol. 23, Issue 6, pp. 661 – 675, June 2004

[51] M. S. Brown, J. G. Goldin, R. D. Suh, M. F. McNitt-Grey, J. W. Sayre and D. R. Aberle, "Lung Microndules: Automated Method for Detection at Thin-Section CT – Initial Experience" Radiology, Vol. 226, pp. 256 – 262, Jan. 2003

[52] K. Marten, A. Grillhösl, T. Seyfarth, S. Obenauer, E. J. Rummeny and C. Engelke, "Computer-assisted detection of pulmonary nodules: evaluation of diagnostic performance using an expert knowledge-based detection system with variable reconstruction slice thickness settings", European Radiology, Vol. 15, No. 2, pp. 203 – 212, Feb. 2005 网络斯兰氏属棕榈酸斯巴丁氏 白化树脂 建冷器 半天地理人的 喜歡解

- a a da a serie de la serie A serie de la serie de la serie dépendént de la serie de la seri A serie de la serie de la serie dépendent de la serie de la ser
- ्रिये के साथ में विकास दिया की मैंसे विविध ने विद्यार प्रसार का सह ते ने तराय तरहर हती है के तरीन स्वाप्त तो का स्वीर्थने स्वाप्त की प्रसार की के का सहय राजनाती के साथ है। से मार्ट के साथ है के से स्वाप्त के से स्वाप् तो जाय के साथ के साथ की साथ की स्वार्थ की स्वार्थ की स्वार्थ की साथ की साथ की स्वार्थ की से स्वाप्त स्वार्थ की

a da antica da servicia da Angelan Sangara Sangara Sangara da Sangara da Sangara Sangara. A servicia da Sangara Sangara Angelanga angelana a servicia da servicia da Sangara Sangara Sangara da Sangara A servicia da servicia da Sangara da Sangara Sangara Sangara da Sangara da Sangara Sangara Sangara Sangara Sang A servicia da servicia da Sangara Sanga

الأمانية المحتوي الأسلامية والمعالمين المحتوية المحتوية المحتوية المحتوية المحتوية المحتوية المحتوية المحتوية ا المحتوية الم المحتوية الم