

UNIVERSITÀ DEGLI STUDI DI BOLOGNA

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI

DOTTORATO DI RICERCA IN FISICA – XVIII CICLO

SETTORE DISCIPLINARE DI AFFERENZA: FIS/07

**A NEW COMPUTER AIDED SYSTEM FOR
THE DETECTION OF NODULES IN LUNG
CT EXAMS**

Tesi di Dottorato di:

ALESSANDRO RICCARDI

Supervisore:

Prof. **RENATO CAMPANINI**

Coordinatore del Dottorato:

Prof. **ROBERTO SOLDATI**

Contents

Abstract.....	iii
Contents	vii
Chapter 1: Lung cancer overview and lung CAD motivation	1
1.1. Some statistics about lung cancer	1
1.2. The question of lung screening	2
1.3. Motivation for CAD systems	3
1.4. Issues concerning CAD systems	4
1.5. Summary	4
1.6. Some examples of lung nodules in CT images	5
Chapter 2: CAD systems in literature.....	9
2.1. Overview of most relevant systems and methods in CAD literature.....	9
2.2. Conclusions.....	14
Chapter 3: Pre-Processing	15
3.1. Processing DICOM images	15
3.2. Segmentation algorithm: motivation	17
3.3. Overview of segmentation algorithm	17
3.4. Image smoothing.....	18
3.5. Histogram thresholding.....	21
3.6. Trachea elimination	22
3.7. Left-Right lung separation	22
3.8. Contour smoothing	23
3.9. Summary	24
Chapter 4: First attempt to build a CAD system.....	27
4.1. Database	27
4.2. General idea about the developing CAD	27
4.3. Nodule enhancing filter: the Matching Filter	28
4.4. Optimizing the Matching Filter threshold	29
4.5. Some cleaning	30

4.6.	Area bounds.....	30
4.7.	Finding borders of signals: the Seeded Region Growing algorithm	31
4.8.	Simple Fast Positive Reduction before classification	31
4.9.	SVM based signals classification and results.....	33
4.10.	Overall system results	36
4.11.	Final note on parameters optimization and results.....	37
4.12.	Critical analysis of the system.....	37
4.13.	Conclusions	38
4.14.	Some examples of processed images	38
Chapter 5: A new CAD, part I: finding nodules in slices		43
5.1.	The Fast Radial transform	43
5.2.	Comparison between Fast Radial and Matching filter	44
5.3.	A new database.....	45
5.4.	A slightly modified Fast Radial transform and other versions of the Fast Radial algorithm.....	46
5.5.	First problem on path: instability of SRG	46
5.6.	A possible solution	47
5.7.	Scale Space.....	48
5.8.	Preliminary tests with Scale Space.....	49
5.9.	Second problem on path and a promising solution	51
5.10.	Putting all together	51
5.11.	Brief summary of Scale Space Approach useful properties.....	53
5.12.	Conclusions	54
Chapter 6: A new CAD, part II: matching signals through slices; coarse False Positive Reduction (FPR-1)		55
6.1.	Grouping 2D signals into 3D objects	55
6.2.	Results of 3D grouping on database DB5mm	56
6.3.	3D objects features	57
6.4.	False Positive Reduction, first part (FPR-1)	59
6.5.	FPR-1 results on DB5mm	60
6.6.	Conclusions	60

Chapter 7: A new CAD, part III: fine False Positive Reduction (FPR-2)	63
7.1. Preparing for signals classification: choosing Regions of Interests and features	63
7.2. Classifying 2D signals: 7-Fold Cross-Validation and 90° rotations of ROIs	65
7.3. Labelling 3D objects	66
7.4. 7-Fold CV results - part I	67
7.5. 7-Fold CV results - part II: choosing positives for training	69
7.6. Introducing Multirotation in training	70
7.7. 7-Fold CV results - part III: effects of Multirotation in training	71
7.8. Introducing Multirotation in test phase	72
7.9. 7-Fold CV results - part IV: effects of Multirotation in test and final FROC curve	73
7.10. False Positives analysis	77
7.11. Conclusions	77
Chapter 8: Analysis of LIDC database	81
8.1. Database at time of writing	81
8.2. CAD processing I: detecting 2D signals	82
8.3. CAD processing II: matching signals through slices; FPR-1	83
8.4. CAD processing III: FPR-2	85
8.5. False Positives analysis	92
8.6. Conclusions	92
Chapter 9: Conclusions	97
Glossary	99
References	101
Acknowledgements	111

Abstract

English version

This work describes the development of a new Computer Aided system (CAD) for the detection of nodules in CT scans of the lung, employing Computer Vision and Pattern Recognition techniques. The system consists of three steps: pre-processing, filtering and False Positive Reduction (FPR). The pre-processing step is dedicated to lung segmentation, and it is mainly based on Gray Level Histogram Thresholding, Seeded Region Growing and Mathematical Morphology. The second and third steps are aimed at detecting nodule-like signals – the filtering step - and at separating these signals into true and false nodules - the FPR step. The main characteristics of the CAD system are: 1) an original and iterative use of the Fast Radial filter, able to detect signals with circular symmetry in CT images; 2) the novel use of a filter based on the Scale-Space theory, able to locate circular signals of a given size; 3) the logical AND of the previous two filters. The iterative application of the Fast Radial filter approximately eliminates one third of the 2D False Positives with respect to the use of a single Fast Radial filter, whilst the Scale-Space based filter cuts 10% to 20% of the 2D False Positives found by the Fast Radial algorithm. The next steps of the system are: 4) a procedure to group signals across adjacent slices, to obtain collections of two dimensional signals corresponding to single 3D candidate nodules, be they true or false ones; 5) a coarse FPR phase, based on length across slices, volume and inclination of 3D candidate nodules, and 6) the fine FPR phase, based on the supervised classifier Support Vector Machine (SVM), fed with Gray Level features extracted from Regions Of Interest located around each signal, whose size and position have been determined by means of the Scale-Space based filter. The system has reached promising results, being able to detect 80% of nodules with 34 FP/Patient, or 65% of nodules with 6 FP/Patient, estimated with a Cross-Validation procedure over 34 nodules of 17 patients, with

diameter between 3 and 10 mm, and with slice thickness 5 mm and reconstruction increment 3 mm.

Final note:

The work described here is both experimental and original. We started off with just a simple and vague problem - find a way to detect nodules in lung CT scans – and realized a fully automated CAD system. Moreover, we have written the thesis with an educational purpose in mind: we have tried to show the pathway that has taken us from the initial problem formulation to the final CAD version, putting special emphasis on the different critical points tackled during the research. This choice has led us to skim rapidly over the theoretical aspects of the already well-known aspects of the employed techniques and to concentrate on the original use of new techniques.

Versione italiana

Il lavoro riguarda lo sviluppo di un nuovo sistema automatico di analisi (CAD: Computer Aided Detection) di immagini TAC del polmone per la ricerca di noduli, possibile segnale dello sviluppo di un tumore. Tale sistema è basato su tecniche di Image Processing e Pattern Recognition. Il sistema è costituito da 3 parti principali: pre-processing, filtering e False Positive Reduction (FPR). Il pre-processing è costituito essenzialmente dall'algoritmo di segmentazione dei polmoni, ed è basato su tecniche di Gray Level Histogram Thresholding, Seeded Region Growing e Mathematical Morphology. La seconda e la terza parte sono invece dedicate a trovare i segnali simili a noduli – la seconda – e a separare questi segnali in veri noduli e falsi noduli – la terza e ultima. I punti salienti del CAD sono dati da: 1) l'adattamento e l'uso iterativo, originale, di un filtro per la rivelazione di segnali a simmetria circolare (Fast Radial); 2) l'uso originale di un secondo filtro in grado di individuare segnali circolari di dimensioni ben definite attraverso un effetto di risonanza (approccio di tipo Scale Space); 3) l'AND logico dei suddetti filtri. L'utilizzo iterato del filtro Fast Radial riduce approssimativamente di un terzo il numero dei segnali Falsi Positivi trovati dal sistema, in confronto al numero dei segnali falsi trovati utilizzando il filtro una sola volta, mentre il filtro di tipo Scale Space riduce il numero dei Falsi Positivi trovati dal filtro Fast Radial di circa il 10% ÷ 20%. I punti successivi del sistema sono: 4) un metodo per raggruppare i segnali che in immagini TAC contigue si trovano in posizioni vicine, in modo che ogni gruppo contenga tutte e sole le immagini di un candidato nodulo; 5) una

prima fase di riduzione dei segnali Falsi Positivi basata su caratteristiche dei raggruppamenti di segnali ottenute dal filtro di tipo Scale Space; 6) una seconda ed ultima fase di riduzione dei segnali Falsi Positivi, basata sul classificatore SVM e su feature di tipo Grey Level ricavate da riquadri costruiti attorno ai segnali in base alle caratteristiche della risonanza fra l'oggetto e il filtro Scale Space. Il sistema ha raggiunto risultati promettenti: 80% di noduli rivelati con 34 Falsi Positivi per Paziente, oppure 65% con 6 FP/Paziente. Tali risultati sono stati ottenuti da una procedura di tipo Cross-Validation su un insieme di 34 noduli di diametro compreso fra 3 e 10 mm, appartenenti a 17 pazienti, in esami TAC ricostruiti a 5 mm di spessore, con 3 mm di reconstruction increment.

Una nota finale.

Questa tesi è la descrizione di un lavoro sperimentale ed originale: partendo da un problema molto semplice e generale, quello di elaborare un metodo per trovare noduli in TAC del polmone, siamo arrivati ad avere un sistema CAD completo. Inoltre abbiamo scelto una modalità di scrittura che seguisse, quanto più possibile, un intento didattico: abbiamo cercato di esporre problemi e soluzioni del percorso che ha portato dalla iniziale formulazione della questione della rivelazione di noduli alla versione finale del sistema CAD. Questa scelta ci ha portato a omettere la descrizione degli aspetti teorici di quegli algoritmi già ben conosciuti, in favore dei nuovi e del loro utilizzo originale.

Chapter 1

Lung cancer overview and lung CAD motivation

1.1 Some statistics about lung cancer

Lung cancer is the leading cause of cancer deaths in Western Countries. In Europe in 2004 there were 1.711.000 cancer deaths. Lung cancer deaths were almost 20% of them (341.800), followed by colorectal (203 700), stomach (137 900) and breast (129 900) [Boyle05]. In Italy there are about 35.000÷40.000 new cases per year, and the incidence is increasing in women [AIRC06].

Tobacco is considered responsible approximately for 85 percent of all cases, while radon exposure for another several percent. Heavy smokers have a much higher risk of dying of lung cancer, about 10 times that of non-smokers [AACR05]. Asbestos, pollution and genetic predisposition are other important risk factors.

Only about 10÷15% of all people who develop lung cancer survive for 5 years, on average, but individual prognosis depends on the extent of disease at the time of diagnosis. Accurate staging, assessing the extent of local and distant disease, is necessary to determine resectability and overall prognosis. Lung cancer staging encompasses 4 stages, from I to IV, with increasing gravity. The survival rate is about 67% if the cancer is detected at stage I, when it is relatively small (no more than 30 mm in diameter), and drops down to less than 1% for stage IV, when metastases have already developed.

1.2 The question of lung screening

The high survival rate of stage I would lead to the conclusion that early detection and treatment would result in a dramatic increase in survival. Unfortunately, lung cancer is usually detected late in the course of the disease, due to the lack of symptoms in its early stages. This is the reason why lung screening programs have been investigated since the '70s, aimed at detecting pulmonary *nodules*: they are small lesions which can be calcified or not, almost spherical in shape or with irregular borders (see Figures at the end of the chapter). The nodule definition for thoracic CT of the Fleischner's Society is "a round opacity, at least moderately well margined and no greater than 3 cm in maximum diameter" [Austin96]. Approximately 40% of lung nodules are malignant, that is, are cancerous: the rest is usually associated with infections. Because malignancy depends on many factors, such as patient age, nodule shape, doubling time, presence of calcification [NLHEP06], after the initial nodule detection further exams are necessary to obtain a diagnosis.

The first kinds of screening protocols were based on chest X-ray (CXR) and sputum cytology, but it resulted that they were relatively insensitive in the detection of small Stage I lung cancers, so they didn't pass into clinical practice. More recently, lung CT has demonstrated to be far superior to X-ray in detecting small and potentially curable nodules [Henschke99]: new generation multislice spiral CT scanners can reconstruct images corresponding to a slice thickness of less than a millimetre, hence are able to clearly visualize objects which are one mm in size [Kalendar00]. At present, only Japan has introduced lung screening in public health care: low-dose spiral CT is annually performed for all smokers aged at least 45. This program has made the 5-year survival rate increase to 58% [Koike99]. In the US and in Europe no lung cancer screening is currently recommended, but this position is being re-examined [Bechtel03], and the debate is far from being over: for example, for Patz and co-workers, who compared patients mortality in a low-dose CT screening group and in a CXR and sputum screening group, screening utility still remains unproven [Patz04]; Swensen and co-workers [Swensen05] say that "our preliminary results do not support *this* possibility (*of reducing mortality*), and may raise concerns that False Positive results and overdiagnosis could actually result in more harm than good"; even, Brenner and co-workers [Brenner04] report an estimated 5% increase in lung cancer incidence among frequently screened smokers. Another important problem to be addressed in the debate

is the specificity of lung CT exams: it is reported that about 20% of suspected nodules reveal to be other mimicking entities or artifacts [Erasmus00]. Moreover, there is a lower size limit for clinically important nodules: some authors suggest that non-calcified nodules less than 5 mm [Henschke04] or 4 mm [Bellomi04] in diameter only justify repetition of annual screening, not immediate work-up.

1.3 Motivation for CAD systems

Besides from the debate about the necessity for lung screening, there are two important facts concerning lung CT analysis that are of great importance:

1. a large inter-reader variability has been reported, indicating the necessity to have at least two or, better, four reads per CT scan [LIDC03];
2. due to the developments in CT technology, the amount of data to be visually inspected by the radiologist is becoming overwhelming.

Both reasons suggest that lung CT CAD systems will become major tools in clinical practice in a few years, even if lung cancer screening will not. The problem of reader variability and oversight in mammography, where CAD is already a well-developed clinical tool, is similar to that in lung CT, and the use of CAD has been demonstrated to be very useful for early detection of breast cancer. Anyway, mammography is the first medical field in which CAD has been applied: in other neighbour fields, such as colon and lung CT inspection, CAD is still in its infancy, and much work needs to be done. At present, there are some recently published studies showing an improvement in radiologists' performance when using CAD: Li and co-workers [Li05b] report an increase from 52% to 68% in the average sensitivity of a group of 14 radiologists in the detection of 17 cancers; Rubin and co-workers [Rubin05] claim an average increase from 50% of single reading to 76% when using their CAD system (SNO-CAD) at a threshold allowing 3 False Positive detections per patient (FP/Patient) on a set of 195 nodules at least 3 mm in diameter; Lee and colleagues [Lee05] compared the detection rates of four radiologists with and without a CAD system (ImageChecker[®] CT LN-1000 by R2 Technology, Inc.) on a group of 78 nodules in the range 4÷15.4 mm in diameter, and concluded that “the CAD system can assist radiologists in detecting pulmonary nodules (...), but with a potential increase in their False Positive rates” and add that “further technological improvements are necessary, in order to increase its sensitivity and specificity”.

1.4 Issues concerning CAD systems

Other than improvements, also general guidelines for CAD assessment, development and use are necessary. Some of them have recently been specified by Gur and colleagues [Gur05]:

1. Cases used for system development and assessment should be pathologically verified;
2. Results should be reported in a standardized manner, and the better way should be the detection rate with a certain number of False Positives per Patient: at present FPs are reported in different units, and this makes difficult to compare different CAD systems.

Other issues are reported in [Goo05], where the author says that it should be important to

1. take into account other underlying diseases in the thorax, which could mislead the CAD system;
2. adjust nodule target range in accordance with task: for example, relevant nodules are in the range 4÷20 mm, because larger ones are not missed by radiologists, and smaller ones are usually considered not relevant [Henschke04][Bellomi04]. Besides, malignancy rarely occurs in nodules less than 5 mm in diameter, but the probability changes in relationship to patient's clinical history;
3. distinguish different types of nodules, i.e. solid, partly solid, and non-solid nodules, which present different clinical significances;
4. characterize nodule malignancy level;
5. integrate CAD systems into Picture Archiving and Communication Systems (PACS).

Another important issue is what a radiologist considers a reasonable amount of False Positives per Patient of a CAD system. Single human readers can go down to 0.1 FP/Patient with detection rates ranging from 40% to 100%, depending on nodule subtleness, and no CAD system is at present barely close to this low number of FPs with a detection rate superior to 50% (see next chapter for an overview of present day CAD systems). The ideal CAD should be able to improve human performances at no or very low cost, and to do that its FP rate should be kept very low - in order not to deceive the

human reader- and this low level should be not far from that of radiologists themselves, which is the situation in mammography CAD [Campanini06].

1.5 Summary

Lung cancer is the leading cause of cancer deaths in the Western world, and many efforts are being undertaken by the scientific community to improve survival chances of patients. Early lung nodule detection is important in this battle, and CT is at present the best medical tool for early nodule detection, even if the debate about lung screening is still open. Anyway, because technological improvements of CT have made image inspection a very demanding task for radiologists, and because of the inherently inter-reader high variability, it is inevitable that CAD systems for lung analysis will be unavoidable tools in medical practice in a few years from now. Before that time, however, many improvements are necessary, as shown by some issues recently highlighted in the scientific community.

1.5 Some examples of lung nodules in CT images

The following figures show some CT images with nodules, whose position have been marked by radiologists with yellow circles.

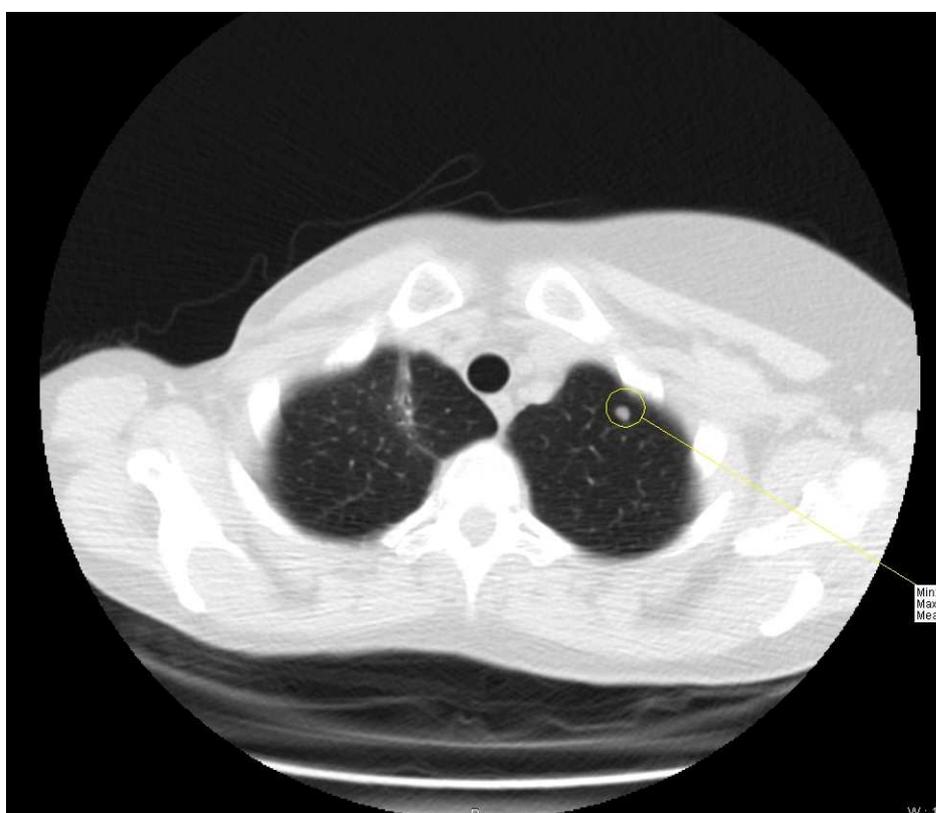


Figure 1: nodule with regular border

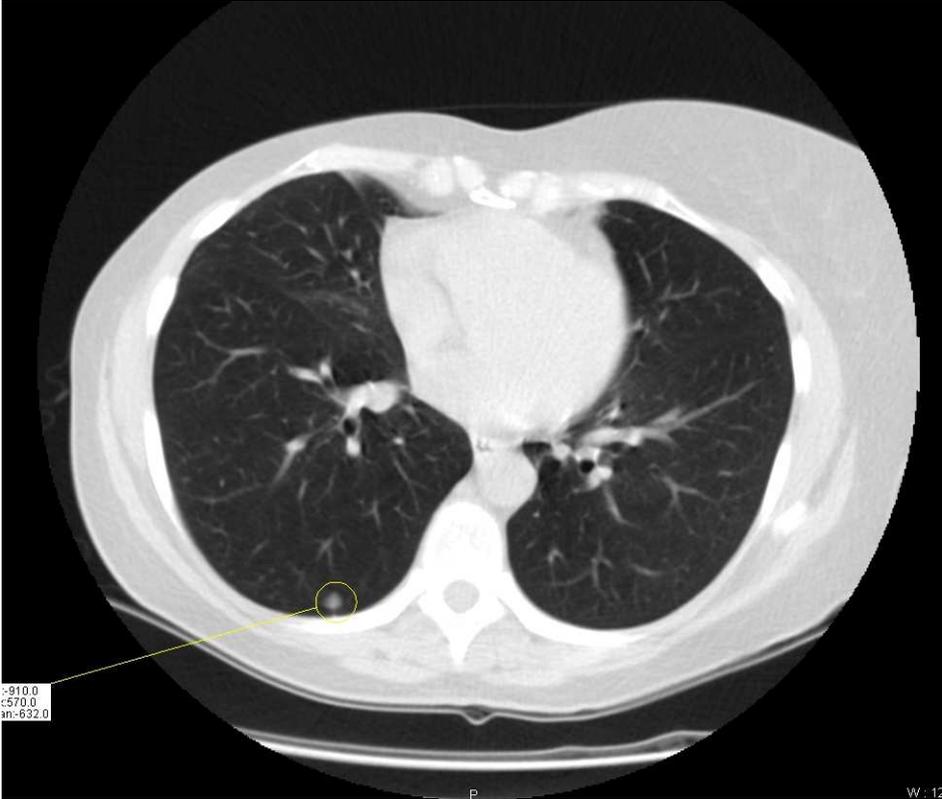


Figure 2: large nodule close to parenchyma

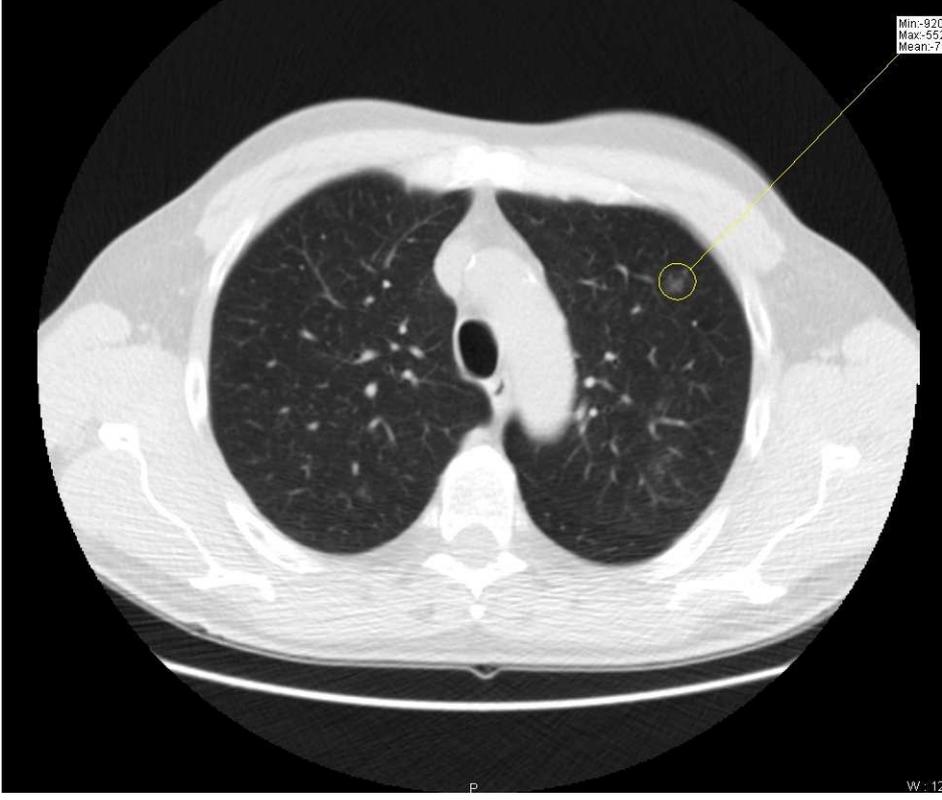


Figure 3: nodule with irregular border

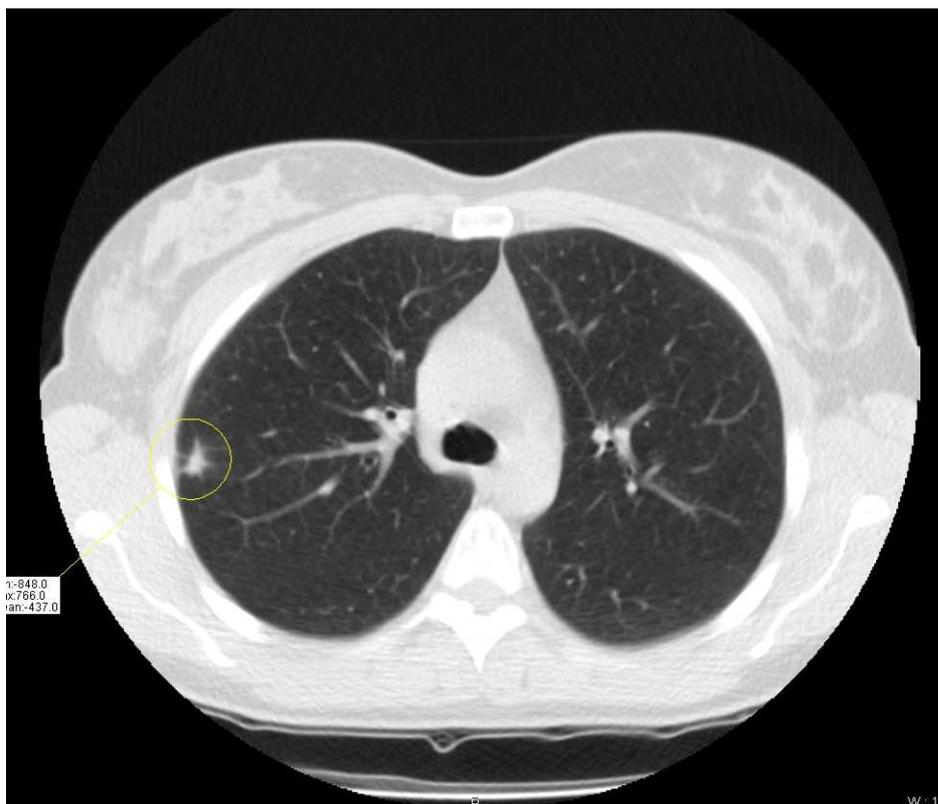


Figure 4: large nodule with irregular border

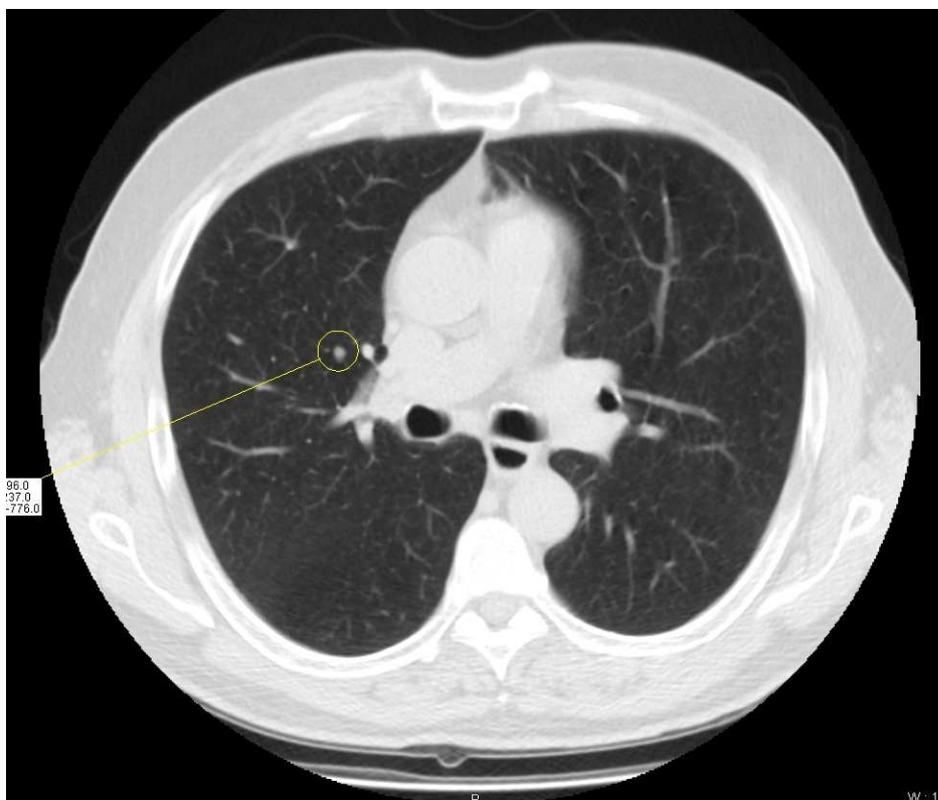


Figure 5: another nodule with regular shape

Chapter 2

CAD systems in literature

Apparently, the first paper about an automated system for the detection of nodules in CT scans of the lung dates back to 1989 [Preteux89]. Since then, a number of approaches have been proposed, especially in recent years, and yet the field can be considered in its youth, with many opportunities of development still waiting to be explored. It should be noted, anyway, that the complete lack of large public databases is one of the reasons that has not facilitated the work of researchers, especially of those that don't have the opportunity to interact with clinicians. Moreover, such databases would very easily allow to compare different algorithms on a common basis. Only recently has such a database, the Lung Image Database Consortium (LIDC), been proposed and its collection begun [Armato04]: at the time of writing it comprises just than 23 cases.

2.1 Overview of most relevant systems and methods in CAD literature

The system presented in [Armato99a] and [-99b] is based on a multiple Grey Level (GL) thresholding of the lung volume: at each threshold, a pixel-connectivity scheme is applied to identify contiguous three-dimensional objects satisfying a volume criterion. After that, a Linear Discriminant Analysis (LDA) classifier is used to distinguish between nodules and non-nodules, represented as vectors of two and three dimensional geometric and GL features of the candidate nodule. The lung volume is previously segmented with a GL histogram thresholding approach, refined with a rolling-ball algorithm. The authors reported an overall sensitivity of 70% with an

average of 3 False Positive (FP) objects per section on a database of 187 nodules coming from 17 patients, using a Leave-One-Out (LOO) procedure (diameters of nodules from 3.1 to 27.8 mm). The slice thickness is 10 mm, as the reconstruction interval. The system, with the major modification introduced by adding a rule-based scheme for a better reduction of candidates prior to classification, is applied to a larger database in [Armato00], whilst in [Armato01] a comparison between scans from low-dose helical CT, with 5 mm slice thickness and 5 mm reconstruction interval, and standard-dose helical CT, with 10 mm slice thickness and reconstruction interval, is presented, showing similar performances of the system on both databases: 71% sensitivity with 1.2-1.5 FPs per section (Nodule median sizes are 6.5 mm and 5.0 mm in the two groups respectively). [Armato02] shows an improvement in performances on low-dose helical CT scans with 10 mm slice thickness and reconstruction interval: 80% sensitivity and 1.0 FPs per slice.

An interesting algorithm for the Reduction of FPs (FPR) is presented in [Suzuki03]: it is named Multi-MTANN (Massive Training Artificial Neural Network) and is based on a Neural Network filtering approach known as Neural Filtering. It is applied to the CAD scheme named before, on a 10 mm thickness and reconstruction interval low-dose database. This kind of Neural Filter is trained to associate to an input image of a candidate nodule an output image representing the distribution for the likelihood of being a nodule. A Gaussian distribution probability is used for nodules, and a constant zero value distribution is used for non-nodules. It shows very good performances on a validation set of 58 nodules: 57 nodules are retained whilst the number of FPs falls approximately from 1 to 0.2 per slice (the overall sensitivity of the system is 80%, the number of FPs per patient changes from 27.4 to 4.8).

In the paper [Lee01] the authors describe a Genetic Algorithm Template Matching (GATM) scheme. They consider the detection of lung nodules as an optimization problem, and use the GA to determine the target position of the candidate nodule in the image, and to select an adequate template image from several reference patterns to consider for the template matching operation. After that, a simple FP Reduction (FPR) step is performed by calculating and thresholding 13 features (geometrical and CT values based ones). The helical CT data set is made of 20 patients with 98 nodules, and the slices are reconstructed at 10 mm interval. The results are a sensitivity of 72% and about 30 FPs per case. Unfortunately, it is not clear if a CV

procedure was considered for the setting of parameters. The authors report also a persistent difficulty in detecting low-contrast nodules, as well as those situated in the apex and in the lung base.

The system presented in [Gurcan02] detects candidate nodules by means of a weighted k -means clustering technique, after an initial step where lung regions are as well segmented within the thorax by a k -means clustering technique applied to the pixel GL values. In the detection step the lung area is subdivided into suspicious and background zones, using a feature vector for the weighted k -means clustering coming from the original image and a median-filtered one. The suspicious areas are then further analyzed: first there are two rule-based FPR steps, exploiting geometrical features in two- and three-dimensions respectively, and then there is a LDA classifier, employing geometrical and GL feature vectors. The overall results, coming from a LOO procedure on a 34 patient database of 63 nodules with mean diameter of 8.9 mm and minimum and maximum of 2.0 and 25.0 mm respectively, with slices thickness of 5, 3 and 2.5 mm, are 84% sensitivity and 1.74 FPs per slice. It is interesting to note that the authors explicitly state that “rule-based systems usually lack the generalization property and should be carefully used in computer vision applications”; however, they say that they attempted to use relatively lax criteria, in order not to be too specific to the used data set, and that these criteria come from radiologists’ experience, hence are expected to be quite well founded. They conclude that further work on a larger data set is necessary to evaluate the robustness of the rule.

An interesting algorithm based on anatomical knowledge of the lungs is described in [Brown03]: a parametric model of the anatomy of the chest is stored in a semantic network, and it is used to guide segmentation - performed by means of attenuation thresholding, region growing and mathematical morphology - and labelling of regions in the lungs. Each node in the network consists of an anatomical name and a set of features, whilst the arcs represent structural relationships and features between anatomical objects. The node features are: X-ray attenuation range in Hounsfield Unit (HU), relative location (“part of”, “inside”), volume, and shape. The system is applied to the detection of nodules (diameter ≥ 3 mm) and micronodules (diameter ≤ 3 mm) in thin-section CT images, with 0.5 – 1.0 mm reconstruction interval. The reported results on a group of 57 nodules (average diameter 6.3 mm) and 22 micronodules (average diameter 2.2 mm) are 100% and 70% sensitivity respectively, with 15 FPs per patient. It

must be noted that only a section of 20 mm longitudinal extent is considered per patient. Another reported result is the improvement in the performances of radiologists, when assisted by the system, from 91% to 95% on nodules and 51% to 74% on micronodules, with an invariant FPs number of 0.3 per case.

In the paper [Yamamoto96] the authors propose a CAD system based on a filter with high sensitivity but low specificity, called Variable N-Quoit, a type of mathematical morphology filter, able to detect over 95% of nodules, together with 400 FPs per patient, which is too a high number. In [Fukano03] this method is given an FPR step, consisting of two different algorithms: the first is a Tophat by Partial Reconstruction filter, used to distinguish blood vessels, and the second is a method that recognizes objects by performing a 3D template matching operation, using artificial templates coming from a ray tracing method applied to nodules blood vessels. These two methods together reduce the number of FPs to 6 per patient (98% reduction), losing only 2 nodules in a set of 37 patients. The total number of nodules of the set is not known. Moreover, it is not clear if the algorithm setting was performed on the same set of the reported results.

The paper [Wiemker02] shows a detection system applied to high resolution CT data of 1 mm slice thickness, with reconstruction interval between 0.5 and 1 mm. In this system, a 2D filter able to identify structures similar to circles or half-circles is applied to binary mask images, which are the result lung segmentation pre-processing. The algorithm is very careful in dealing with nodules attached to the pleura. A three-dimensional shape analysis carried out on binary threshold images is the core of the FPR step. A particular relevance is given by the authors to the usefulness of the average Hounsfield value in rejecting FPs. The direct utilization of the Hounsfield values is due to the absence of the partial volume effect, for objects greater than 1 mm, in this case. The reported results are 95 % sensitivity and 4.4 FPs per patient, on a database of 203 nodules larger than 2 mm in diameter. It is not clear if this set is the same used for the setting of the algorithm parameters.

The core of the CAD system given in [Kaucic03] is a mathematical model that describes the formation of the CT image of the lung by taking into account the physics of the scanner and the anatomy of the chest. During the process of detection of the structures within the lungs, a Bayesian statistical framework is employed to select the model that provides the most probable interpretation of the data. The probability of each

model is calculated from image intensities, geometric features and neighbourhood properties. Results on a 2.5 mm collimation low-dose CT database of 50 cases with 43 nodules are 70% detection rate and 8.3 FPs per case. An average improvement of 3.9% in the detection rate in a group of 3 radiologists is also reported. From the paper it is not clear if the test set was different from the one used to set the algorithm parameters.

Ge and co-workers have introduced a method for false positive reduction based on a Linear Discriminant Classifier employing features drawn from 3D gradient field and 3D ellipsoid fitting [Ge05]. Purpose of the gradient features is to discriminate objects with approximately radial GL distribution from objects with highly asymmetrical distributions. Fitting parameters of the 3D ellipsoids are also given to the classifier. Reported results of a Leave-One-Out patient scheme are: 80% sensitivity, over a group of 111 nodules with size 3÷30 mm, with 0.34 FP/slice. The database consisted of 82 CT scan (3551 slices) from 56 patients, with slice thickness 1.0÷2.5 mm, so a figure of 14.7 FP/scan (or 21.6 FP/Patient) can be inferred.

A filter with the ability to selectively enhance dot and lines is shown in [Li03]: it is based on the properties of the Hessian matrix of dots and lines, taken as models of nodules and vessels, which are the targets of the study performed on two- and three-dimensional thoracic CT images. This is probably the first attempt to create a pre-processing filter with both good sensitivity and specificity for the targeted object. The method has been reported to be effective for nodule enhancement and false positive reduction [Li05a].

Paik and co-workers [Paik04] describe a new type of approach to lung nodules and colonic polyps detection in helical CT based on the Surface Normal Overlap (SNO) algorithm. They propose to use SNO as the first step of a larger detection scheme, in conjunction with a subsequent FPR phase. SNO detection is based on the calculation of a scoring for objects which have survived the pre-processing and segmentation phase. This scoring is proportional to the number of surface normals that pass through or near the object. Convex objects with dominant curvature along two directions, like nodule and polyps, tend to have larger scoring than vessels, which have a dominant curvature along one direction only. The authors optimized the algorithm on simulated CT data, and reported promising results on a set of 8 CT scans. Interesting results are also described in [Rubin05], where a SNO-CAD has been used to compare performances of radiologists with and without lung CAD.

In the paper [Armato03], the authors compare the results obtained by the CAD system explained in [Armato99a] on two databases: one with patient projection reconstructed with the “standard” algorithm, and the other with the same patient projection data reconstructed with the “lung” algorithm. The comparison demonstrates similar levels of performance, which makes the author claim for an encouraging degree of robustness of their system.

2.2 Conclusions

Probably not much more than a couple of dozen of approaches have been proposed in the last 15 years to realize a CAD system for lung nodule detection in CT scans, so we can say that most of the work still needs to be done: since its beginning, the CAD community has been mainly focused on systems for mammography, which are just becoming clinical tools after many years of development, and it is only now really turning on lung, so we can expect an explosion of research in the next 5 years. At present there is only one commercial system available (ImageChecker® by R2 Technology, Inc.), and many others are being developed. The lack of a public and common database, necessary for validation and comparison, has really dampened the work of researchers in this field, but the recently proposed LIDC database should overcome this crucial problem.

Chapter 3

Pre-processing

Pre-processing is usually a very important step of CAD systems, though a very general and vague subject, which depends on the peculiarities of the system and of the tackled problem: in this CAD it comprises the transformation of images from the DICOM format, which is the general standard used in medicine to store digital data, into Gray Level images, and the segmentation of the inner part of the lungs, to be subsequently analysed by the CAD core.

3.1 Processing DICOM images

The Digital Imaging and Communications in Medicine (DICOM) standard was created by the National Electrical Manufacturers Association (NEMA) to aid the distribution and viewing of medical images, such as CT scans, MRI, and Ultrasound [DICOM06]. Part 10 of the standard describes a file format for the distribution of images. All lung CT scans used in our studies are exactly in this format.

A single DICOM file contains both a header (which stores information about patient, type of scan, image dimensions, etc), as well as all of the image data. In our case image data is saved as 16bit/pixel signed values. For different reasons (compression, etc) lung CT image data stored in DICOM files are in raw format and have no particular meaning (units of measure), hence to evaluate a CT scan a quantitative scale exists that describes relative attenuations: the Hounsfield scale. This scale assigns water an attenuation value (Hounsfield Units, or HU) of zero. The range of CT numbers is 2,000 HU wide although some modern scanners have a greater range of

HU up to 4,000. Each number represents a shade of gray with +1,000 (white) and –1,000 (black) at either end of spectrum (see Table 1).

BONE	+400 +1,000
SOFT TISSUE	+40 +80
WATER	0
FAT	-60 –100
LUNG	-400 –600
AIR	-1,000

Table 1: Hounsfield Scale of CT numbers.

To convert raw CT data into Hounsfield Units a linear transformation is applied: $HU = \text{Raw Pixel Value} * \text{Rescale Slope} + \text{Rescale Intercept}$. This transform is applied to each pixel. The values for Rescale Slope and Rescale Intercept are stored in the DICOM image header.

Another pixel value transformation is needed in order to work with CT images: Window Level and Window Width Transform. Whilst the range of CT numbers recognized by the computer is 2,000, the human eye cannot accurately distinguish between 2,000 different shades of gray. Therefore to allow the observer to interpret the image, only a limited number of HU are displayed. A useful gray scale is achieved by setting the Window Level and Window Width on the computer console to a suitable range of Hounsfield Units, depending on the tissue being studied. The term Window Level represents the central Hounsfield Unit of all the numbers within the Window Width. The Window Width covers the HU of all the tissues of interest which are displayed as various shades of gray. Tissues with CT numbers outside this range are displayed as either black or white. For example, when performing a CT examination of the chest, a WW of 350 and WL of +40 are chosen to image the mediastinum (soft tissue), whilst an optimal WW of 1500 and WL of –600 are used to assess the lung fields (mostly air): see Figure 1 below.



Figure 1: These two images are of the same section, viewed at different window settings. (Left) A window level of +40 with a window width of 350 reveals structures within the mediastinum but no lung parenchyma can be seen. (Right) The window level is -600 with a window width of 1500 Hounsfield units. This enables details of the lung parenchyma to be seen, at the expense of the mediastinum.

3.2 Segmentation algorithm: motivation

Nodules are located within the lungs, in an area which is usually no more than half of the area of the CT slice: this means that a lot of processing time can be saved if the searching algorithms only run on this inner part. Moreover, the number of False Positives found in a segmented image is dramatically lower than that found in the same image without segmentation, because no signals at all will be found outside of the lungs. Indeed, lung segmentation is a common pre-processing step of many CAD systems [Armato99a][Gurcan02] and in general of lung image analysis systems [Hu01][Brown97]. Given the large size of CT datasets, manually segmenting the lungs is tedious and prone to inter observer variations. That is why we implemented and tested a fully automatic segmentation algorithm for lung CT volume data.

3.3 Overview of segmentation algorithm

Our segmentation algorithm is composed of five main steps. First a smoothing algorithm is applied to the CT stack to reduce noise. Then the lung region is extracted from the CT images by adaptive Gray Level thresholding. Afterwards trachea region is eliminated from initial CT slices. The left and right lungs are then separated to permit finer processing on each lung separately. Finally, to include nodules attached to the lung wall the lung contour is smoothed. The overall segmentation process is described in Figure 2 together with the type of data involved in each processing step.

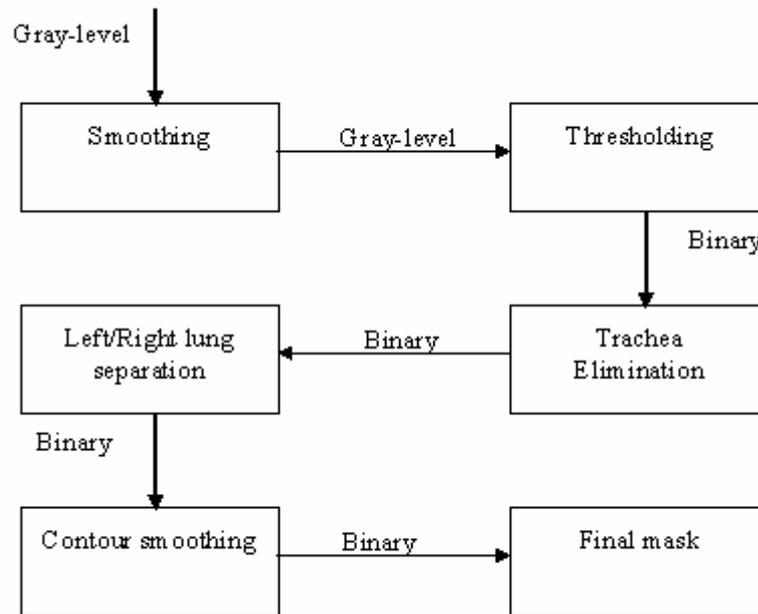


Figure 2: Overall segmentation algorithm. Arrows show direction of data flow.

3.4 Image smoothing

In this step a smoothing algorithm is applied in order to reduce noise and prepare images for segmentation. Because order-statistics filters are known i) to provide excellent noise-reduction capabilities and ii) to introduce less blurring than linear smoothing filters of the same size [Gonzalez02][Huang79], we used a median filter. We tested 3x3 box, 5x5 box, 7x7 box median filters. Examples are given in Figures 3-6. We have chosen to use the 3x3 box median filter because it cleans images very well and at the same time it has a good edge-preserving ability: larger boxes have too heavy a blurring effect on object borders and details.

It must be noted that smoothing is performed only for segmentation: the detection system, described in the subsequent chapters, is applied on the original, unsmoothed, images.



Figure 3: original image.



Figure 4: result of 3x3 median filter application on image of Figure 3: noise has been very well suppressed, but edges have been preserved.



Figure 5: result of 5x5 median filter application on image of Figure 3: blurring begins to be evident, many details have been lost.



Figure 6: result of 7x7 median filter application on image of Figure 3: blurring is heavy.

3.5 Histogram thresholding

During this step the actual binary masks for the lung area are determined. Binary masks are generated from input gray-level CT data using an iterative thresholding algorithm [Ridler78], a better method than the conventional thresholding algorithm, in which the threshold is simply chosen as the minimum between the two maxima of the GL histogram (see Figure 7 below) [Castleman96]. The image histogram is initially divided into two parts using a starting threshold value, which can be for example half the maximum of the dynamic range of the current image, or the conventional threshold value just described. Afterwards, the sample mean of the gray values associated with the foreground pixels and the sample mean of the gray values associated with the background pixels are computed, and a new threshold value is determined as the average of these two sample means. The process is repeated until the threshold value does not change any more (the algorithm has been proved to converge by its author).

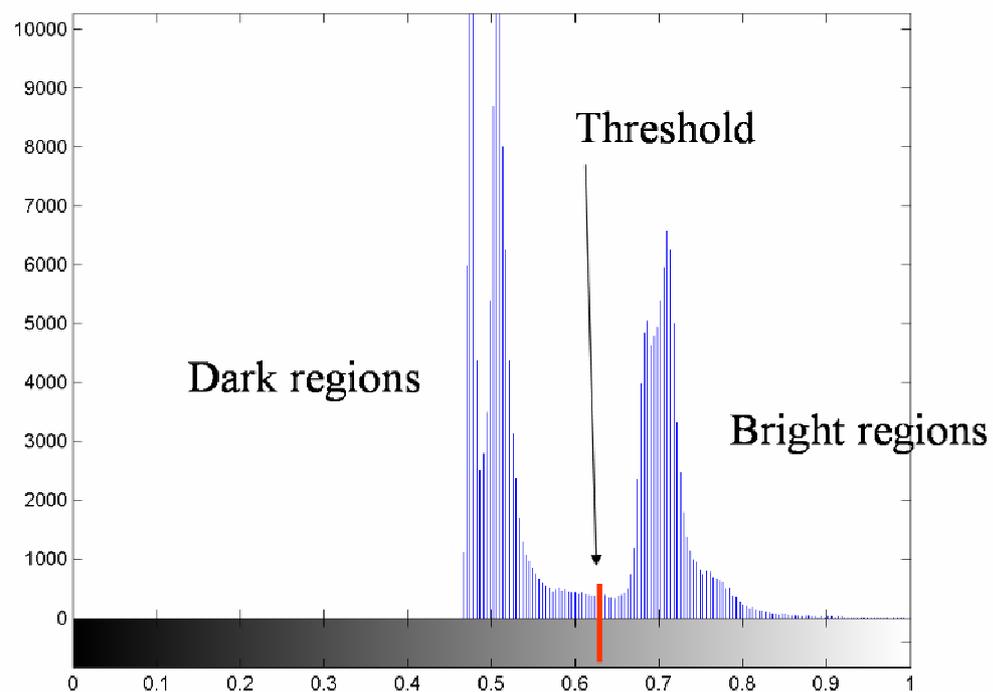


Figure 7: example of Grey Level histogram of a CT slice (GL values have been normalized between 0.0 and 1.0). The histogram clearly shows that most of the pixels of the image belong to either a dark region (the background), or a bright region (the foreground). The threshold can be used to divide these two zones in the image (histogram thresholding).

Because lung walls and ribs in CT scans are brighter than any other lung region, further processing is necessary to obtain masks containing only lung parenchyma. After thresholding, morphological binary opening [Soille99] is applied to each mask to

remove very small regions (a few pixels), usually caused by noise. The structuring element is a disk of radius 3. Then binary masks are inverted and structures that are lighter than their surroundings and that are connected to the image border are suppressed. Finally holes inside the lung region are filled. All these steps can be seen in Figure 8.

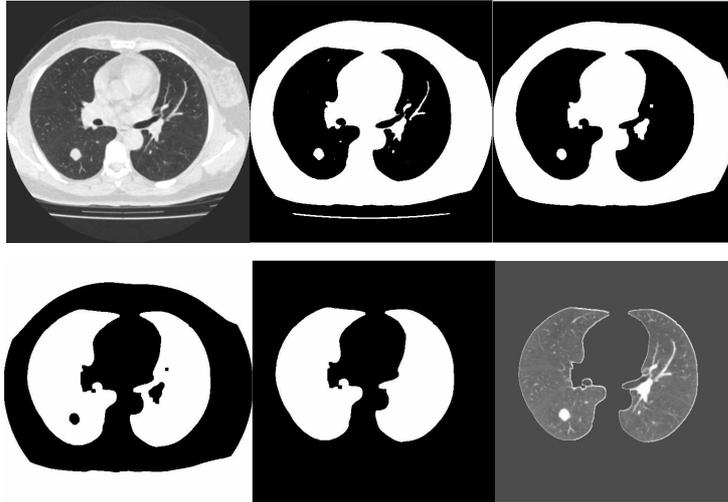


Figure 8: Thresholding step is made up of smaller processing sub-steps: (from left to right, top to bottom) Single slice before thresholding, Binary mask after thresholding, Binary mask after morphological opening, Binary mask after inversion, Binary mask after border regions elimination and hole filling, Final masked image.

3.6 Trachea elimination

Initial lung segmentation based on gray-level thresholding tends to include the trachea. To ensure that this structure does not contribute to the segmented lung regions, trachea is eliminated from the segmented thorax region through Seeded Region Growing (SRG) algorithm. (SRG has also been used in a stage of the CAD development. For some more information on it and critical notes see paragraph 7 and 12, chapter 4 and paragraph 5, chapter 5) Seed points for trachea segmentation are automatically identified in the superior CT sections, because the trachea is always the only segmented object in these slices. Seeded Region Growing is used to expand the identified seed points into the corresponding trachea regions. Region growing ceases when an area-based criterion is satisfied. Seed points are identified in subsequent sections based on the center-of-mass location of the segmented trachea region in the previous section. The result of trachea elimination step can be seen in Figure 9.

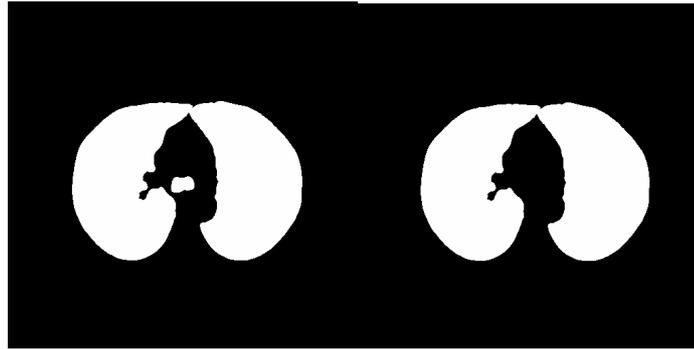


Figure 9: Lung segmentation before (left) and after (right) trachea elimination.

3.7 Left-Right lung separation

The presence of a single large segmented region in any section indicates that the two lung regions are “fused” at the anterior junction. Distinction between left and right lungs is often required for more detailed analysis. Consequently, the single lung region is separated into two regions by eliminating pixels along the anterior junction line. The separation into left and right lung is carried out on the “accumulated” image obtained by summing all binary masks obtained so far along the z-axis. The Grey Level “accumulated” image thus obtained is thresholded using successfully increasing threshold values until the resulting binary image contains two distinct lung regions. Results from these steps can be seen in Figure 10.

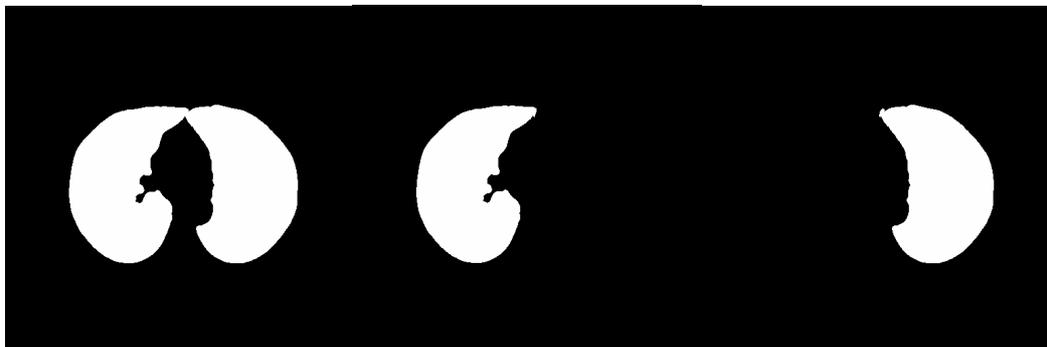


Figure 10: Left/Right Lung separation. Lungs attached at the anterior junction (left figure) are separated into left and right lung (centre and right figures).

3.8 Contour smoothing

In some cases, pulmonary nodules are adjacent to the pleura and have similar densities. Therefore, when Grey Level thresholding step is applied, concave depressions appear at the nodule location in the lung mask. In order to obtain the correct extraction of the pulmonary parenchyma and not to lose the nodule, this depression must be rebuilt. Figure 11 shows examples of an image with a nodule adjacent to the pleura and its mask with the depression resulting from applying the previously described steps. The

pulmonary parenchyma is reconstructed through morphological binary closing with a disk structuring element of radius equal to the radius of the biggest nodule we are looking for (20 mm in diameter: see also chapter 1, paragraph 4). Results from this step can be seen in the two images at the bottom of Figure 11.

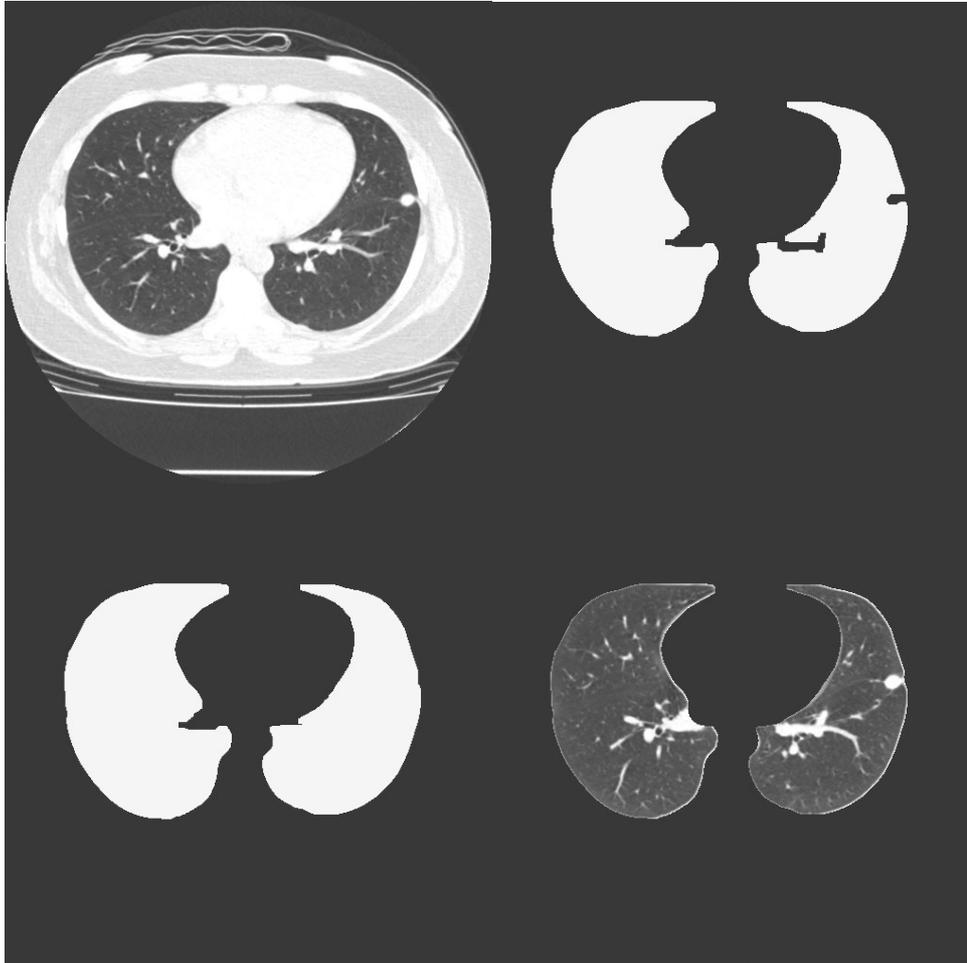


Figure 11: Application of contour smoothing. Original image (top, left); Concave depression at nodule location in the outer border, and large depression due to a bright area close to the inner border (top, right); Rebuilt depressions in binary mask (bottom, left); Final masked image of the lungs (bottom, right).

3.9 Summary

Pre-processing of the CAD system described in this thesis includes transformation of CT data from DICOM to Gray Level images and lung segmentation. Segmentation is a 5-step process mainly based on GL Histogram thresholding and morphological filtering; trachea elimination and left and right lungs separations are accomplished through SRG and an *ad hoc* “accumulation” process respectively. The image is median-filtered before the whole 5 step process, in order to eliminate as much as possible the effect of noise. The segmentation algorithm described here is simple

though effective. Nonetheless, many possible improvements could be tested in future: for example, different and more advanced smoothing filters, contour tracing algorithms, etc.

Chapter 4

First attempt to build a CAD system

It is a rule of life, when facing a problem, after studying what other researchers have already found, to try to solve it by oneself, and better is to forget a little bit their results. This was no exception. To get in touch with the problem, it was decided to start with a simple system, and then to analyze its behaviour and its pros and cons, and use them as a guide to develop a new, and hopefully better, CAD system.

4.1 Database

The database (DB) considered in this chapter (DB10mm) is composed of 22 patients with 34 nodules, with diameters between 4.1 mm and 17.4 mm (average 10 mm), and the average number of slices per patient is 65. All the scans come from a Low Dose Spiral CT (GE HiSpeed CT/i), with parameters: 140 kVp, 40 mA, slice thickness: 10 mm, reconstruction increment: 5 mm, slice dimensions: 512 x 512 pixels (0.67 ÷ 0.74 mm/pixel). All the image values have been transformed into 8 bit code, i.e. 256 Grey Levels. The database comes from IEO, Istituto Europeo di Oncologia, Milan, Italy, where a team of radiologists detected and marked the nodules.

4.2 General idea about the developing CAD

The general idea of the CAD algorithm to be developed at this stage was the following: after the segmentation of the inner part of the lungs, consider a procedure to locate the candidate nodules in each CT slice, then calculate some geometric features and use them to train a classifier, hopefully able to discriminate between true and false

2D signals, that is, true nodules and false nodules in each CT image, then mark the surviving objects on the CT image, to be subsequently inspected by radiologists. This is a 2D method, that is, it analyzes the CT scans image by image and does not take into account any relationship between signals in consecutive images.

4.3 Nodule enhancing filter: the Matching Filter

Because lung nodules in CT images are most of the time bright objects, nearly circular in shape, situated in a noisy but darker background, together with other bright objects, the simplest of the shape detector, the Matching Filter (MF), was chosen as signal detector [Castleman96]. The MF (aka Matched Detector) is an algorithm which allows to calculate the correlation between a signal, for example an image $f(x,y)$ containing objects and regions, and a template signal, a spatially smaller image $h(x,y)$, whose presence in the image we want to locate. If there is a match, the correlation of the two functions will be at its maximum where h finds a correspondence in f . By properly thresholding the correlation image, these peaks are found and the template-like signals located. As a nodule template it was chosen a small square image with varying side lengths: 3, 5, 7, 9 and 11 pixels. The Grey Level values were arranged in order to mimic a typical nodule profile, almost circular in shape: see Table 1 as an example.

135	150	150	150	135
150	150	155	150	150
150	155	155	155	150
150	150	155	150	150
135	150	150	150	135

Table 1: 5 by 5 nodule template, Grey Level values (8-bit)

It should be noted that a different intensity of the filter values, with the same relative ratio, it is of no relevance with respect to the enhanced shapes, affecting only the tuning of the threshold at which signals are detected. If the GL values of the filter are left unchanged during the processing, only the threshold must be optimized. The values in Table 1, at about 3/5 of the maximum GL, were chosen for convenience. The chosen filter was 5 by 5 pixels in size. It was found indeed that the smaller the template the finer the signal localization and the smaller the size of the enhanced objects: a 3 by 3 template is too sensitive to noisy objects, and a 7 by 7 one, or wider, leaves the filtered image too much blurred, a fact that makes difficult to choose a proper threshold. The

choice was done by looking at the quality of the filtered images for the subsequent thresholding step.

4.4 Optimizing the Matching Filter threshold

Optimizing the threshold is a very important step in the detection process. A high threshold would give less signals, with great chance to loose true signals; a low threshold, on the contrary, would give too many signals, most of which false ones, or, even worse, many signals would be merged together (see Figures 4, 5, 6 at the end of the chapter). In order to set the optimal threshold, the histogram of the correlation image was analyzed (see Figure 1 as an example).

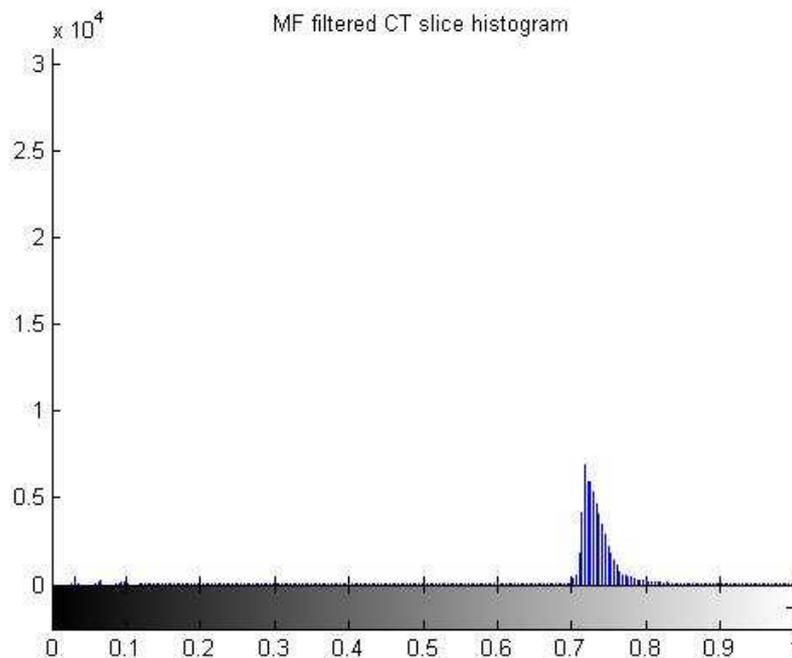


Figure 1: example of GL histogram of a MF processed CT slice. The peak at zero is not taken into account for the subsequent processing.

It can be noted that only one peak is always present (the zero GL peak on the left is ignored), and, moreover, its corresponding GL value is always lower than that of the nodules that might be in the image, and the ratio between these two values can vary between 1 and 10. This was verified on a small subsample of nodules of the database. It was then decided to set the optimal threshold in the following way:

$$threshold = peak_GL + numdelta * delta_value$$

where *numdelta* is a parameter to be optimized on the database, and $delta_value = (255 - peak_GL) / 255 * delta_number$, that is, the difference between the peak GL and the maximum GL divided into a predefined number of parts, *delta_number*, set to 10 in this

case, normalized by the maximum GL, 255 in the case of 8-bit images. It is clear at this point that the threshold varies with the image according to the position of the GL peak, but the relationship between this peak and the threshold is determined by the *numdelta* parameter, to be established by means of a validation database and never changed for images of the same database. *Numdelta* was set to 1.65.

4.5 Some cleaning

The thresholded image, now a binary one, was then processed with an erosion filter [Castleman96] of side 3 by 3 pixels (see Table 2), with the aim to:

1. Eliminate objects with area of 1 or 2 pixels
2. Clean large objects borders, very irregular in shape
3. Part very large and elongated objects into smaller and more manageable ones.

0	1	0
1	1	1
0	1	0

Table 2: erosion filter, diamond shaped

In a typical image, the number of objects after the erosion decreases from 50/150 to 15/70, and most of the deleted signals are 1 or 2 pixels in area. The filter was set 3 by 3 because a “heavier” erosion by a larger filter would have deleted too many pixels. The diamond shape proved to delete less objects than the square one, so it was chosen for precaution.

4.6 Area bounds

The subsequent processing consists in an area thresholding: because nodules are limited in size, and because most of the remaining signals smaller than 3 pixels are false signals, all the objects with area larger than MAXAREA and smaller than MINAREA are deleted. MAXAREA was set to 430 pixels, far larger than the area of the larger of the nodules at this point of processing, which was about 250 pixels: unfortunately the DB was too small to allow a normal Cross-Validation procedure, so it was decided to proceed by setting the threshold-type parameters on the whole DB, but in a “light” manner, that is, not too strict, so not to risk too much an overfitting bias (see also paragraph 4.11 at the end of the chapter).

At this point, detected signals are given a positive label (they are considered nodules) if their centre of mass is within a certain tolerance from the centre of mass of

the nodules marked by the radiologist, otherwise it is given a negative label (it is a false signal). Tolerance was set to 6 pixels, with Euclidean distance.

4.7 Finding borders of signals: the Seeded Region Growing algorithm

Now the problem was to find the exact borders of the objects. To do this, a Seeded Region Growing (SRG) procedure was implemented [Mehnert97]. SRG is a robust and rapid algorithm for objects segmentation in intensity images that, taking as input a number of seeds, either individual pixels or regions, is able to make the seeds “grow” until the number of free pixel around the grown region is zero, or a certain stopping criterion is satisfied. For each seed, pixels are examined around it and if, for example, their GL value is within a certain tolerance from the mean GL of the seed, they are merged with the seed, otherwise are discarded. The merging criterion can be updated or not at each step. The stopping criterion can be introduced at will.

The SRG procedure was then used to make the signals that had survived the preceding detection steps grow. Pixels connected to the seed were merged only if their GL value was within 5% of the mean GL of the seed: no noticeable changes were found over a number of objects in the range between 3% and 9% of the mean GL. It was also noted that objects growing too much were highly likely to be not nodules, but maybe vessels or other lung structures. Thus, another stopping criterion was introduced: a maximum number of iterations in the growing process. At any iteration all the pixels which are neighbours of the seed region or of the already grown one are considered for merging, and, if merged, their neighbours are put in the list of new neighbours to be analyzed at the next iteration. When the number of iterations for each region becomes too big, it means that the object is growing too much, and can be stopped. By analyzing the distribution of iterations in many patients, it was noted that:

1. most of the nodules grow thoroughly within 100 iterations;
2. the mean number of iterations in an image is typically 1,000 – 10,000;
3. the maximum number of iterations can go up to 40,000.

For this reasons the iteration number was bound to 200.

4.8 Simple False Positive Reduction before Classification

At this stage of development, making the system run over the entire DB gave the following results: 28 of 34 nodules found with 54 2D signals, 16.3 2D False Positive signals (FP) per image, on average. In order to try to eliminate some of the FP before

using the classifier, some simple geometric characteristics of the signals were considered for analysis. These characteristics were: area, major axis length, minor axis length, ratio of major and minor axis lengths. Figures 2 and 3 illustrate these characteristics, comparing nodules and non-nodules. To optimize as much as possible the thresholds on these values, half of the patients of the database were randomly chosen and analyzed: because the database is small, in order to avoid overfitting bias not all the patients were considered at this stage.

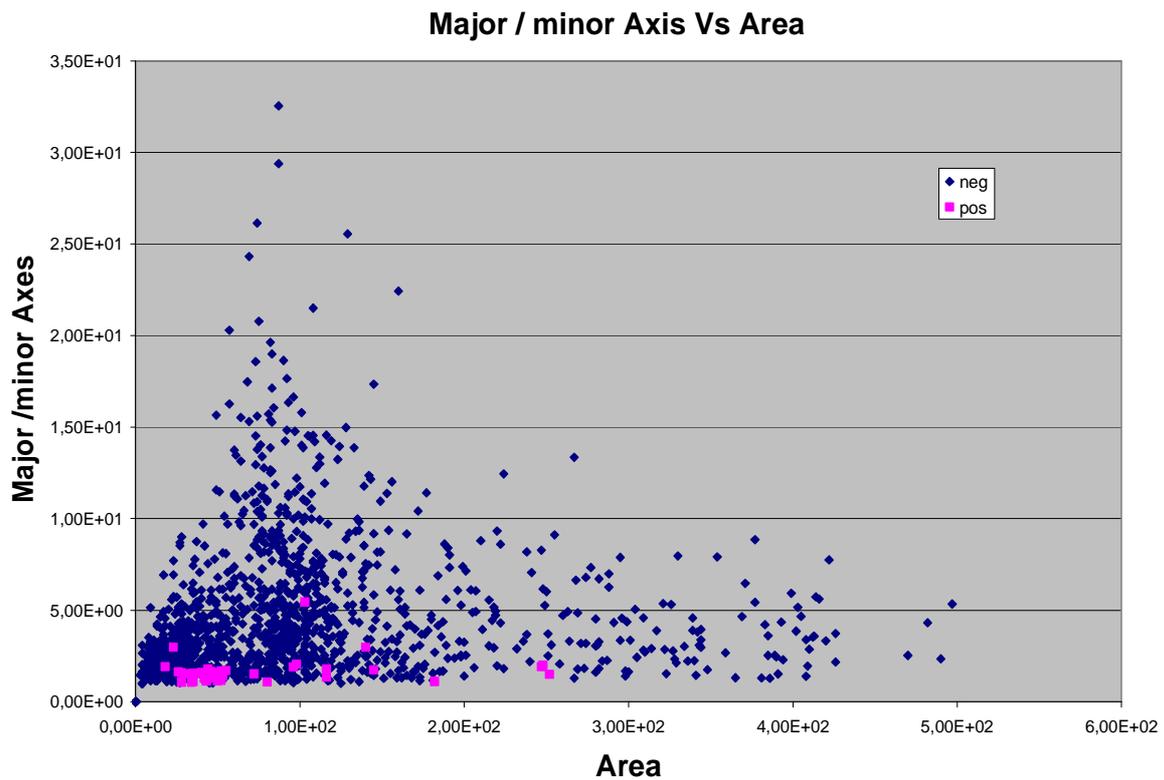


Figure 2: Major/minor axis length ratio Vs Area. Blue: negative signals (FP), lilac: positive signals (nodules).

Only thresholds on the simple characteristics were considered, and the chosen values are the following ones:

1. area < 310 pixels;
2. Major axis length < 40 pixels;
3. minor axis length < 25 pixels;
4. ratio Major/minor axis length < 4.5pixels.

It can be noted that some of the two dimensional nodules get lost in this simple False Positive Reduction (FPR) step, but this is inevitable, when trying to diminish the overall number of FP. However, there is some redundancy here: there are indeed 51 2D true

signals remaining of the 54 initial ones, and none of the 28 found nodules are lost (Note that a nodule is considered detected when at least one of its corresponding 2D views has been found). On the other hand, the simple FPR step proved to be very effective, because about 35% of the FP signals are eliminated: now there are only 10.6 FP per image on average.

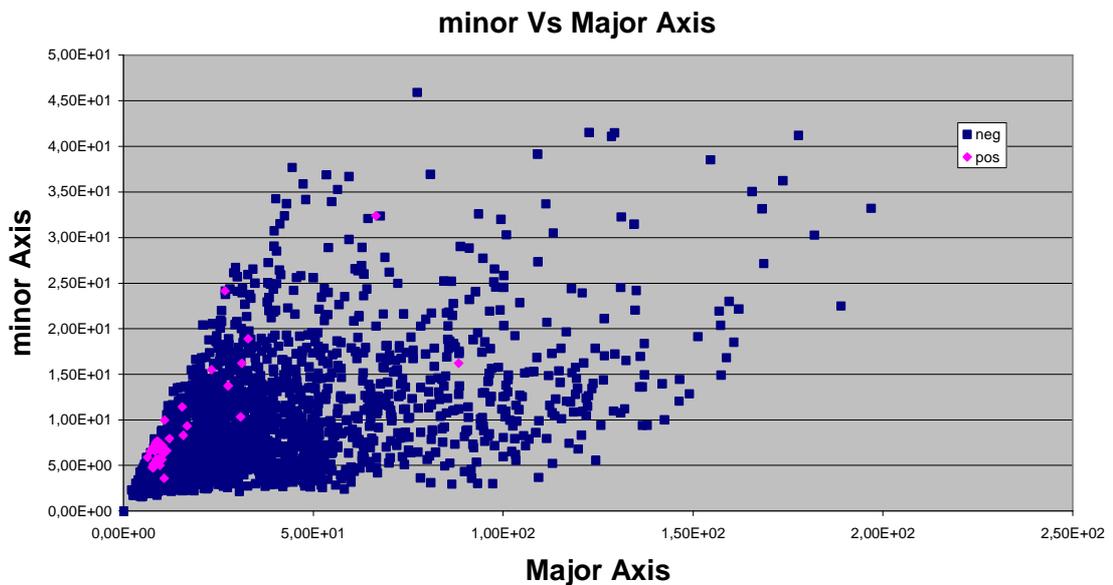


Figure 3: Major axis length Vs minor axis length. Blue: negative signals (FP); lilac: positive signals (nodules).

4.9 SVM based signals classification and results

At this point, with 51 2D positive signals and 15,444 2D negative signals coming from the first part of the detection algorithm, a Support Vector Machine (SVM) classifier [Vapnik95] was used to discriminate between nodules and not-nodules. SVM is a class of supervised learning algorithms for nonlinear classification, regression, distribution estimation, novelty detection and clustering. It is deep-rooted in the Statistical Learning Theory [Vapnik98] and at present is considered the state-of-art in the machine learning field: indeed, it exhibits many useful properties that are not shared, together or at all, by other algorithms. Among these properties are: 1) the training process is made by solving a quadratic optimization problem, hence no local minima trapping is ever encountered; 2) the architecture of the classifier is automatically set by the solution of the optimization problem, which makes the training very easy, and, moreover, allows for a very low risk of poor generalization; 3) the generalization capability is good even when a small number of training samples are employed or when the two classes are unbalanced to a high degree (the number n of training samples is

considered *small* when d , the sample space dimensionality, is not much smaller than n [Vapnik95], even if the seriousness of the situation also depends on the absolute value of d , due to the *curse of dimensionality* [Haykin99]). These properties guarantee in general a learning procedure simpler and faster than that given, for example, by a conventional multilayer Artificial Neural Network (ANN), or by a Radial Basis Function (RBF).

Perimeter, area, major axis length, minor axis length, equivalent diameter, circularity, compactness, mean Grey Level and standard deviation of Grey Level are the features calculated for each signal. Circularity and compactness are defined as

$$circularity = \frac{object \cap same_area_circle_in_same_center_of_mass}{object},$$

$$compactness = \frac{4\pi * area}{perimeter^2} : \text{they are two ways to measure how much an object shape}$$

resembles a circle. Note that geometric features are invariant to rotation and translation transformations, but not to scaling ones.

Because there are not enough data to set up a standard Estimation, Validation and Test procedure [Haykin99], a Cross Validation (CV) algorithm [Efron83] was instead employed, in particular, a *stratified 10-Fold CV* procedure, putting an equal number of samples of each class in 10 subsets, and then training on the different groupings of 9 subsets and testing on the 10th subset each time, for 10 times. When the number of samples is small, it is particularly important to optimize the use of the database. This goal implies that both the *inner* (result of learning optimization procedure) and the *outer* (result of learning supervisor decisions) parameters of the learning algorithm are optimized: the SVM algorithm is able to extract a great amount of information from the training set, automatically used to optimize the inner parameters, but the optimization of the outer parameters depends on the global training procedure, on the relationship between the training and the test sets, generally speaking. For example, it is well-known that the Hold-Out method is far less efficient than the CV [Kohavi95] for training any algorithm and giving an estimate of its performances in case of a small database. Efficiency can be measured in terms of reliability of results at varying number of samples: a plot of this behaviour gives what is called a *learning curve* [Hastie01]. The faster the plateau (or saturation level) is reached, the more efficient the algorithm is. When the number of samples is very big, almost all

algorithms show the same results, but when their number is not very large, which is what usually happens in the real world, efficiency heavily comes into play. The 10-Fold CV algorithm is a very good one for giving an estimate of the performances of a learning algorithm in small database cases. It can be affected by bias and variance to some reasonable degree, and it is considered one of the best choices [Hastie01].

Another possible trick that can be used in cases when the two classes are very much unbalanced is to randomly decrease the number of negative signals to a figure always larger than the positive signals number. This can improve the speed of training, without affecting too much the final results, providing the class is not under-sampled.

The SVM parameters that were optimized with the CV procedure were: the kernel, the Cost parameter C , the unbalancing Cost parameter $C^{+/-}$. Used kernels were the Linear, the Polynomial of degree 2, 3 and 4, the RBF with γ equal to 0.01, 0.05, 0.1, 0.5, 1.0, 2.0, 5.0, 10.0. The parameter C was given the values 0.05, 0.1, 0.5, 1.0, 2.0, 4.0, 8.0, 10.0, 12.0, 15.0, 20.0, and the parameter C^{\pm} the value equal to the ratio *number of positive signals/number of negative signals* in the training subset. The parameter $C^{+/-}$, changing the cost of one class with respect to the other in the training phase, moves the final output decision line, which affects the label given to objects [Riccardi00]. The number of negatives used in training was sampled at 60, 100, 150, 200, 300, 400, 500, 600, 700, 800, 1000, 2000. The database underwent a *whitening* normalization process [Duda00][Campanini04] which simply forces the samples distribution to have zero mean and unit standard deviation: this is necessary in order to have a training procedure where all the dimensions of the samples are given the same weight, hence are treated for their intrinsic discriminatory capability and not for their original range values [Riccardi00][Bazzani01].

A summary of results are given in Table 3: for each kernel, only the best ones are shown.

Kernel	Kernel parameter	C	Negatives number in training	Hit Positives	Hit Negatives
Linear	-	10÷20	> 2000 (13%)	(85.4±0.2)%	(80.1±0.1)%
Poly	2	0.5÷2	> 1000 (6.5%)	(82.3±0.3)%	(79.2±0.2)%
RBF	0.5	0.05÷1	> 1000 (6.5%)	(80.4±0.4)%	(80.7±0.2)%

Table 3: summary of CV results of SVM classifier trained on the 51 positive signals and 15,444 negative signals.

Performance percentages and errors are calculated as mean and standard deviation of the mean of 100 Monte Carlo runs: in each run a random redistribution of samples is executed before the CV is performed. Though this procedure cannot give an estimate of the variance of the CV procedure [Kohavi95][Efron83], nonetheless it can provide us with an estimate of the variance in the database at hand, hence can approximately tell us how far the system is from the plateau in the learning curve. The small errors show that the system has almost reached the saturation level in the learning curve, at least for what concerns the given database; it is indeed not possible to check if this is true also with respect to the positives and negatives global distribution, because this distribution is unknown.

The Linear kernel shows the best results, the number of negatives used for training can be a small percentage of the total number of 15,444, the value of the cost parameter C can be chosen in a small range. The classifier shows very good results, as it is able to correctly find about 85% of positives, in a population of only 51 samples, and 80% of negatives, in a population of more than 15,000.

4.10 Overall system results

Now it is possible to give an estimate of the overall performances of this first CAD system. The number of detected nodules before the SVM classifier was 28/34 (0.8235), with 10.6 False Positives per Image (FP/Image): these figures, combined with the SVM classifier results, lead to $0.8235 * 0.854 = 0.7033$ nodule fraction, and $10.6 * 0.199 = 2.1$ FP/Image. In other words, approximately 70% of nodules have been detected, together with about 2 false signals per CT slice. It must be noted that the redundancy of 2D nodule signals with respect to the 34 nodules has not been considered, so 70% could be considered a lower limit. The final figure of system performances is then 70% of nodules @ 2.1 FP/Image. These results can be considered comparable with those of a similar system developed by Armato and co-workers [Armato99a], even if our system shows a larger variability because of the smaller database.

4.11 Final note on parameters optimization and results

A final note about parameters optimization procedures and results of this work is necessary. Due to the small database at disposal, many of the parameters of the CAD system, apart from the SVM classifier, have not undergone a complete validation procedure, that is, have not been tested on an independent validation set. They have been chosen according to their behaviour on a subset of the database, as much as possible a well subsampled one, and in a very conservative manner: they could have left less false signals than what they actually did, but we decided to stay far from the limits. This cautiousness should help avoiding as much as possible the overfitting risk that is always present when using small databases, even though it is not possible to give a final figure of the overall performance error of the system without a complete validation procedure. However, an estimate of the error can be given using the binomial proportion interval estimation [Berger95][Brown02]: considering 24 found nodules over 34, at 95% Confidence Interval the error is approximately 5÷6 nodules.

Anyway, it must be stressed that the goal of this part of the research was primarily to get in touch with the problem, to obtain an estimate of the behaviour of a very simple approach to nodule detection.

4.12 Critical analysis of the system

A critical analysis of the CAD system leads to the following notes:

1. the well known drawbacks of the MF have been verified in this work, in particular: 1) high GL pixel regions are enhanced even if their shapes do not match with the template, 2) the enhanced signal is spread over an area wider than the signal itself [Rao94]. As a consequence, the system finds three kinds of FP: nodule-like, bright zones close to the border, circular vessels (see Figure 7 at the end of the chapter);
2. the geometric features have some positive properties: they are translation and rotation invariant, and a negative property: they strongly depend on the object border, which in turn depends on the SRG algorithm. Unfortunately, the SRG algorithm can be unstable, especially when the seed is very small [Adams94]. Moreover, the geometric features do not take into account the context of the signal, and its texture;

3. the detection algorithm is 2D-only, and this is another limitation: more discriminative information could be found considering the relationships between signals across scans;
4. a final limitation is due to the kind of data, in particular their thickness, limited to 10 mm: this means that nodules smaller than 10 mm are detected with much difficulty.

4.13 Conclusions

This first system has provided us with a lot of information which hopefully will be helpful in developing a more advanced CAD system. In particular we know that we have to: i) find a filter with high sensitivity *and* high specificity to circular signals of the size corresponding to nodules, ii) take into account the 3D behaviour of the detected signals across slices, iii) find classification features which are good for discriminating positive and false signals and, as much as possible, not dependent on the unstable SRG algorithm, iv) base the analysis on a large database composed of CT scans with slices thinner than 10 mm, in order to be able to detect signals in the size range clinically relevant (4÷20 mm, see paragraph 1, chapter 4).

4.14 Some examples of processed images

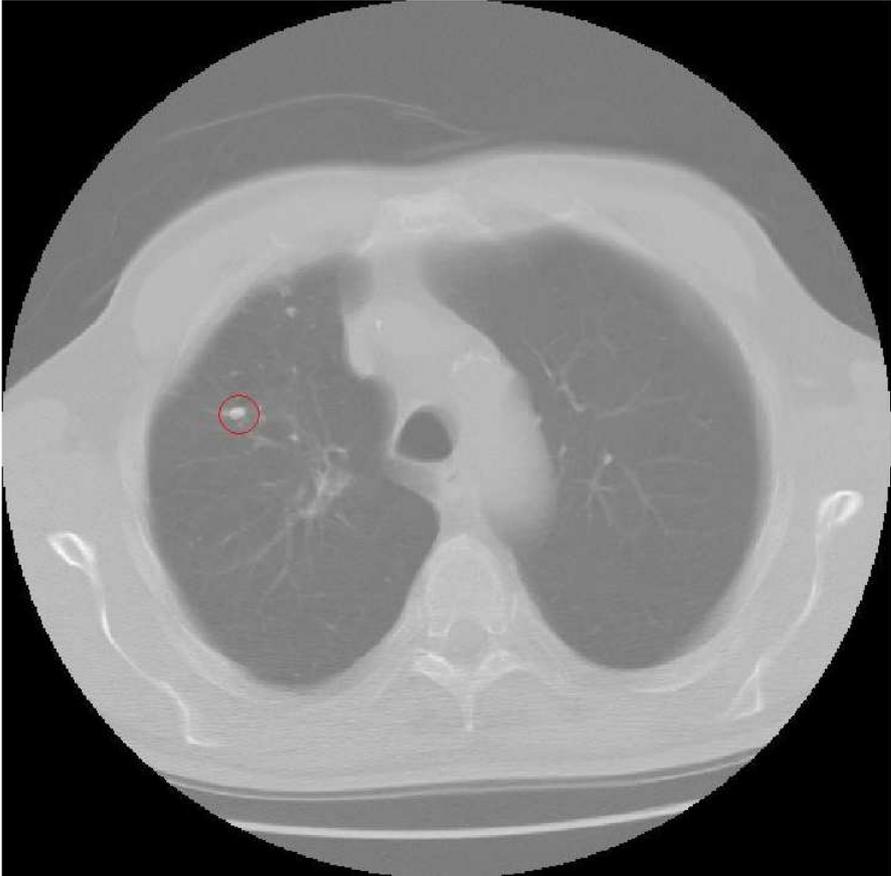


Figure 4: example of CT slice with a nodule marked by radiologists (red circle).

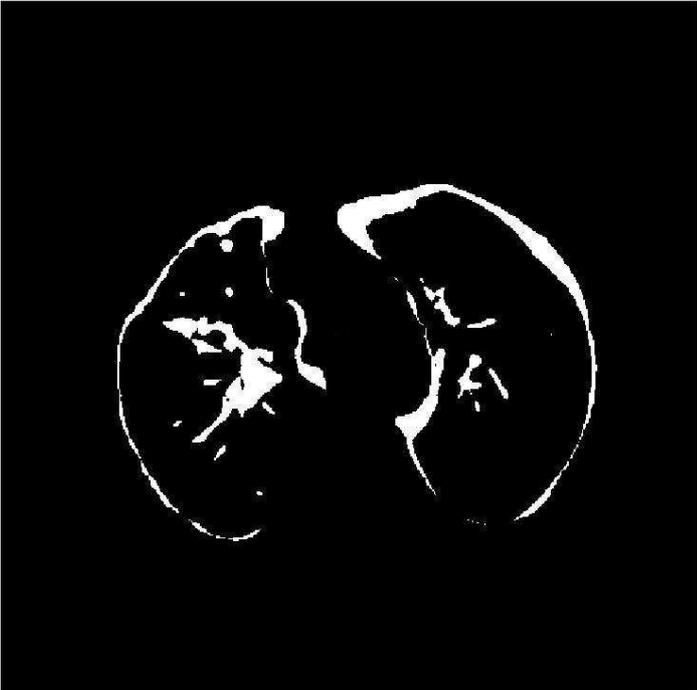


Figure 5: slice of Figure 4 after MF filtering and thresholding. The threshold is too low, hence many large zones are enhanced.

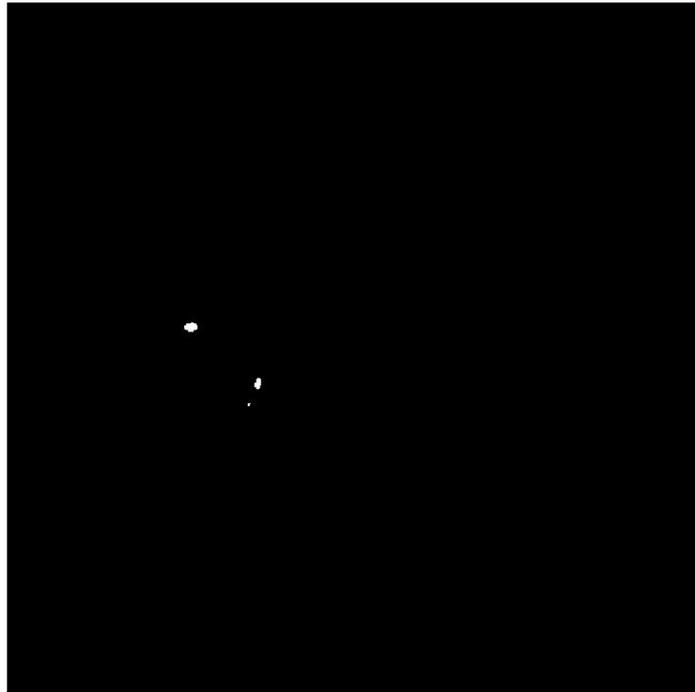


Figure 6: slice of Figure 4 after MF filtering and thresholding. Threshold is high, hence a small number of objects have survived.

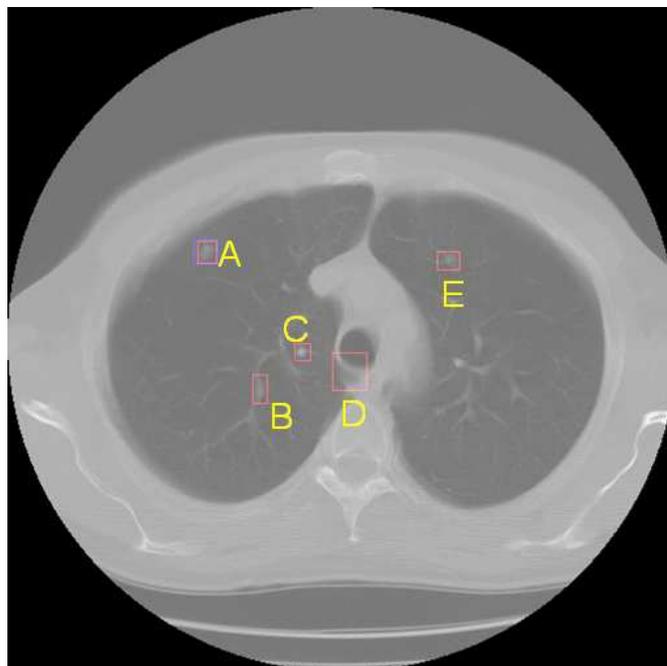


Figure 7: examples of objects detected by the CAD system. Together with a nodule (blue and red squares, object A), 4 false nodules have been found (only red squares, objects B, C, D, E). Object C is an example of a nodule-like false signal, as well as B and E, object D is one of the false signals due to the enhancement given by the MF to high Grey Level zones close to the segmentation border.

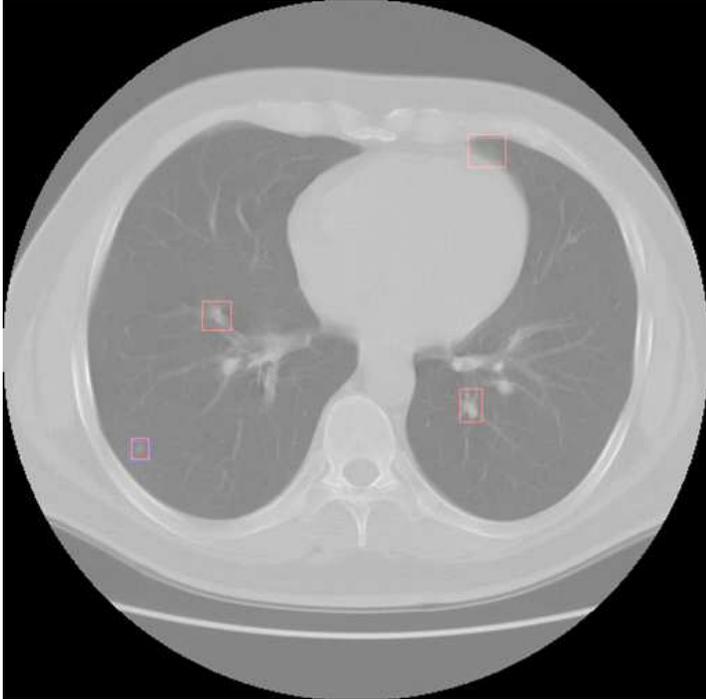


Figure 8: another example of detected objects.

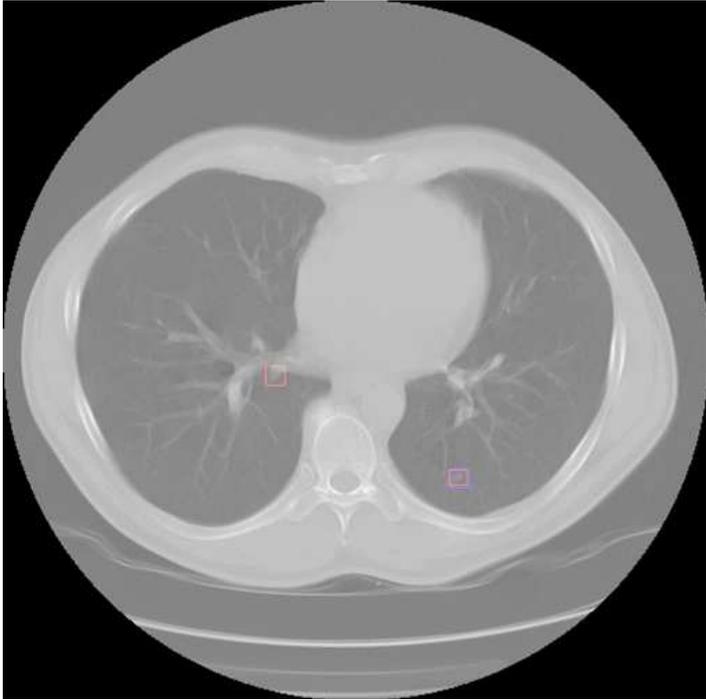


Figure 9: another slice with more examples of detected objects.

Chapter 5

A new CAD, part I: finding nodules in slices

The observations about the weaknesses of the MF-based CAD described in the previous chapter constituted the basis for developing a new, and improved, CAD system. In this chapter, in particular, we will focus on a different and more specific method to find nodule-like signals in (2D) CT slices.

5.1 The Fast Radial transform

Since most nodules in CT images usually show a quasi-circular shape, and the few of them more irregular look nonetheless like compact bright “stains” in a darker background, a filter able to enhance bright zone with radial symmetry would be a natural choice. A very interesting filter for radial symmetry has recently been presented by Loy and Zelinsky [Loy03]: the Fast Radial Symmetry transform (FR). This transform has been shown to provide equal or superior performance when compared to contemporary techniques such as Reisfield’s Generalized Symmetry Transform [Reisfield95], Di Gesù and Valenti’s Discrete Symmetry Transform [Di Gesù95], Kovesi’s symmetry from Local Phase [Kovesi97].

The FR transform determines the contribution each pixel \mathbf{p} makes to the symmetry of pixels around it, at an integer distance (radius) of n pixels from the above mentioned central pixel \mathbf{p} : the value of the transform at radius n indicates the contribution to radial symmetry of the pixels lying in a circle around it at distance n . At each radius n an *orientation image* O_n and a *magnitude projection image* M_n are generated by examining the gradient \mathbf{g} at each point \mathbf{p} . The point \mathbf{p} influences a

positive-affected point \mathbf{p}_+ , which is the point the gradient is pointing to, and a negative-affected point \mathbf{p}_- , the point the gradient is pointing away from (see Figure 1).

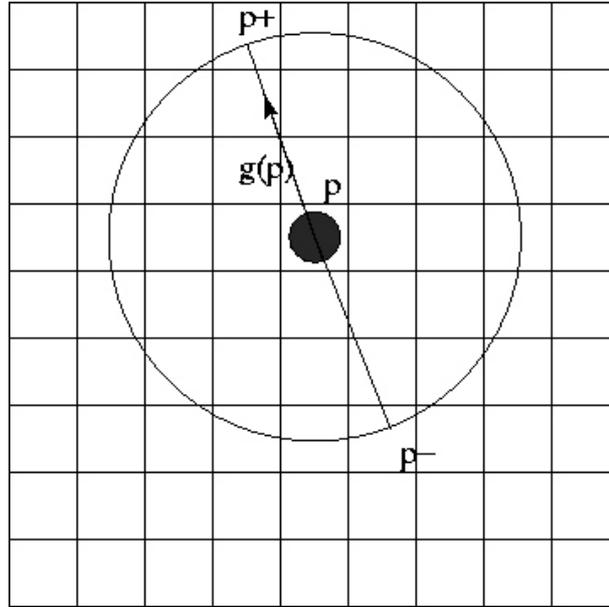


Figure 1: positive-affected and negative-affected points by gradient $\mathbf{g}(\mathbf{p})$ at point \mathbf{p} . Radius $n = 3$.

The projection images are initially zero, and are then built in this way:

$$\begin{aligned} O_n(\mathbf{p}_+(\mathbf{p})) &= O_n(\mathbf{p}_+(\mathbf{p})) + 1, \\ O_n(\mathbf{p}_-(\mathbf{p})) &= O_n(\mathbf{p}_-(\mathbf{p})) - 1, \\ M_n(\mathbf{p}_+(\mathbf{p})) &= M_n(\mathbf{p}_+(\mathbf{p})) + \|\mathbf{g}(\mathbf{p})\|, \\ M_n(\mathbf{p}_-(\mathbf{p})) &= M_n(\mathbf{p}_-(\mathbf{p})) - \|\mathbf{g}(\mathbf{p})\|. \end{aligned}$$

Afterwards, the radial symmetry contribution at radius n is calculated as $S_n = F_n * A_n$,

where $F_n(\mathbf{p}) = \frac{M_n(\mathbf{p})}{k_n} \left(\frac{|\tilde{O}_n(\mathbf{p})|}{k_n} \right)^\alpha$ and $\tilde{O}_n(p) = \{O_n(p) \text{ if } O_n(p) < k_n, k_n \text{ otherwise}\}$. A_n

is a two-dimensional Gaussian used to spread the influence of affected pixels as a function of radius n , k_n is a scaling factor used to normalize M_n and O_n across different radii, and α is the parameter that controls the radial strictness. The full transform is then

defined as $S = \frac{1}{N} \sum_n S_n$, and its positive and negative values correspond respectively to

bright and dark radially symmetric regions.

5.2 Comparison between Fast Radial and Matching Filter

FR parameter k_n was set as suggested by authors. To find the optimum radius range some CT images with and without nodules were analyzed: it was found that $n > 9$

can enhance many falses, together with nodule-like objects, then it was decided to use together radius values equal to 3, 5, 7, 9 pixels, approximately corresponding to nodule sizes in the range 4÷10 mm. This does not affect the detection of larger nodules, whose bulks are always enhanced, but it can save from detecting too many large false signals. The strictness parameter α was set to 2: a larger value asks for too strict circular objects, so many irregular nodules are lost, whereas value 1 is not specific enough, and finds too many falses. This is also the value for parameter α suggested by authors.

The FR filter was initially compared with the MF on the whole database DB10mm: results are given in Table 1.

<i>filter</i>	<i>MF optimal THR</i>	<i>FRTHR 0.4</i>	<i>FRTHR 0.48</i>	<i>FRTHR 0.52</i>
<i>Hit- nodules</i>	28/34	31/34	30/34	29/34
<i>Hit – 2D signals</i>	51	63	58	56
<i>FP/Slice before cuts</i>	16,3	16,6	12,5	10,9

Table 1: results of comparison between MF and FR on DB10mm. *Hit nodules* refers to individual nodules, *hit 2D signals* to 2D signals that compose a nodule. Clearly, the FR algorithm is a better detector than the MF algorithm.

For the MF it was considered the optimal threshold value (THR) described in chapter 4, while for FR were analyzed thresholds from 0.3 to 0.6, with 0.02 increments. These tests have clearly shown that FR is better than MF for the analyzed task. At the rate of about 16 FP/image, FR is able to find 63 2D signals instead of the 51 of MF (approximately 25% increase), whereas at the rate of 10÷11 FP/image, the increase is 5 signals, but without any FPR step, like that described in paragraph 8, chapter 4, for MF. Hence, not only less FPs are detected with the same, or a better, True Positives detection rate with respect to the MF, but also the FR is easier to be optimized: the filtered image is simply thresholded without the need for other manipulations, necessary instead after the MF algorithm processing (see paragraphs 4, 5 and 6 in chapter 4). The variations on the nodule detection rates are not significant because their number is too small.

5.3 A new database

A new database has been used in the rest of this work (DB5mm): it is composed of 13 patients with 34 nodules, with diameters between 2.7 mm and 9.7 mm (22 nodules are larger than 4.0 mm and 34 than 3.0 mm), plus 4 patients without nodules. The

average number of slices per patient is 105. All the scans come from a Low Dose Spiral CT (GE HiSpeed CT/i), with parameters: 140 kVp, 80 mA. Slice thickness is 5 mm, reconstruction increment is 3 mm, slice dimensions are 512 x 512 pixels (0.67 ÷ 0.74 mm/pixel), and reconstruction filter are *lung*, and *standard*. The database comes from IEO, Istituto Europeo di Oncologia, Milan, Italy, where a team of radiologists detected and marked the nodules. Tolerance was set to 6 pixels, with Euclidean distance.

5.4 A slightly modified Fast Radial transform and other versions of the Fast Radial algorithm

A slightly modified version of the FR algorithm [Kovesi05b] is introduced here. In this version, the matrix images M_n and O_n are not zeroed at the beginning of each n -value radius loop: the consequence is that more weight is given to small search radii in the full transform S , when more than one radius is considered. By performing many comparisons between the original FR, the slightly modified FR, the *orientation-based* FR (ORBA) [Loy03] and the *noise-lowered* FR algorithms (NLO) [Loy03], we have found that *the slightly modified FR leads to the best results*, for nodule detection, hence it has been chosen for the CAD.

In the *noise-lowered* FR, gradients smaller than a pre-defined threshold β are ignored, because more corrupted by noise. In the *orientation-based* FR, it is the *magnitude projection image* M_n that is ignored, together with small gradients. We tested these algorithms on the DB5mm database, with the noise parameter β equal to 0, 0.05, 0.1, 0.15, 0.20. Always has the above-mentioned modified FR algorithm given the best performances: with $\beta \geq 0.15$ both ORBA and NLO didn't get over 55 2D nodule hits with 35 FP/image (FR: 88 hits), and with $\beta \leq 0.10$ both reached 80 hits at no less than 45 FP/image (FR: 89 hits at 38 FP/image). From this point, the modified FR has been employed, and simply referred to as Fast Radial.

5.5 First problem on path: instability of SRG

At this point of work, we had to face a big question about how to proceed. We started by performing several tests in order to analyze the impact of seed area on SRG and consequently on the calculation of geometric features of nodule-like signals detected by the FR algorithm after thresholding. We shrank to a different extent the detected signals before starting SRG, and then compared the ratio between major and minor axes length of grown signals. It must be noted that some erosion, or shrinking, is

necessary before making the signals grow, because it can happen that the signal masks in the binary thresholded image are larger than original signals. We found that *the shape of positive signals becomes less and less regular with increasing degree of shrinking*: for example, the maximum ratio of axes goes from 3.1 when just one pixel is eroded, to 4.4 when shrinking leaves just one pixel. This behaviour is due to the instability of the SRG algorithm at the diminishing seed area [Adams94]. This instability affects the determination of the borders of the objects, which are fundamental to calculate geometric features to be used for object classification: if borders are unstable, geometric features are unreliable and classification results are unreliable. Unfortunately, this is only the first kind of problem introduced by the pair SRG-geometric features: indeed, there is another problem. When it comes to consider 3D objects in CT slices, that is, follow 2D signals in consecutive slices and group them together to characterize 3D nodules as collections of 2D signals, it might happen that SRG grown 2D objects are hard to put together, due to the strange shape they can acquire during the growing process – growing too much or too less – which is strictly related to the given seed and the stopping conditions chosen. Furthermore, objects can grow one into another, adding more troubles, if possible. For example, when it happens that very strange 2D objects, extremely irregular in shape, need to be grouped across slices, what do we have to consider as coordinates of the objects, to be compared? Would it be enough to take the centre of mass? After this operation, to what extent would the 3D object shape be reliable and ready for further processing? There is no easy answer to these questions.

5.6 A possible solution

The most pleasant choice would probably be to get rid of any kind of shape algorithms, hence of geometric features: this, in turn, means that we have to find an alternative way to determine the general shape of the objects, and a very effective way to extract information from their texture. Well, we might be lucky. We actually know we don't really need to use geometric features, because we have access to many classification algorithms based on different kinds of features, such as the multiresolution analysis methods – wavelets, ranklets [Masotti05], etc –or the Support Vector Regression (SVR) Filtering [Angelini05], to name but a few, that have already proved to be very effective as classification features. So, one part of our problem could have been solved. In fact, a different problem arises from this prospected solution: all these methods are based on the determination of a proper Region of Interest (ROI)

around the object to be analyzed. In other words, we have to know the dimensional extent of the signal, in order to tailor the best ROI around it. Apparently, the SRG comes again into play, with all its drawbacks. Well, only apparently, in the end. It turned out that the problem of a tailored ROI can be very well, and very appealingly, solved by a Scale Space based Approach (SSA) [Lindeberg93].

5.7 Scale Space

Scale Space representation was introduced by Witkin [Witkin83] and Koenderink [Koenderink84], and provides a good framework for dealing with image structures at different scales. Scale Space of an image $I(x, y)$ can be defined as a function $L(x, y, \sigma)$ produced by the convolution of a variable-scale Gaussian $G(x, y, \sigma)$ with the image:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left[-(x^2 + y^2)/2\sigma^2\right]$$

Unfortunately, the original approach does not address the problem of how to select appropriate scale for further analysis, as stated by Lindeberg [Lindeberg93], who proposed an heuristic principle to deal with this problem: “*a scale level at which some, possibly non-linear, combination of normalized derivatives assumes a local maximum can be treated as reflecting the characteristic length of a corresponding structure in the data*”. The basic idea of the principle stems from the behaviour of normalized Gaussian derivatives of sinusoidal signals: it can be shown that the scale (sigma) at which a normalized derivative assumes its maximum is proportional to the wavelength of the signal. Normalized (dimensionless) coordinates are defined as: $\xi = x/\sigma$, hence the normalized derivative operator is: $\partial_\xi = \sigma \partial_x$. The Normalized Laplacian (NL) thus becomes: $\sigma^2 \nabla^2$. According to the principle, Scale Space extrema - points that are local extrema both in scale and space - of the NL reflect a characteristic length of the objects in the image. In particular, it turns out that the ratio between the scale levels at which the maxima of analysed objects, similar in shape but different in size, are attained, is roughly equal to the ratio between the sizes of the objects.

This approach is basically the core of the SIFT algorithm [Lowe04], even if it employs Difference of Gaussians (DOG) filters instead of the Normalized Laplacian: they can be shown to be equivalent, by using the diffusion equation:

$$\sigma \nabla^2 G = \frac{\partial G}{\partial \sigma} \cong \frac{G(x, y, k\sigma) - G(x, y, \sigma)}{k\sigma - \sigma},$$

which gives: $G(x, y, k\sigma) - G(x, y, \sigma) \cong (k-1)\sigma^2 \nabla^2 G$, that is, the Difference of two Gaussian filters with sigma equal to $k\sigma$ and σ is approximately equivalent to a Normalized Laplacian multiplied by $(k-1)$, where k is integer. Above all, DOG are faster in execution than NL.

5.8 Preliminary tests with Scale Space

We performed some experiments with the SSA, using both NL and DOG filters, and analyzed the relationship between the sigma value of local maxima and the sizes of simple bright objects, like circles, squares and rectangles, in a dark background. In the case of the DOG approach, the algorithm initially computes the correlation of the original image with a Gaussian filter with increasing sigma (the implementation is actually done by iteratively filtering the already filtered image: this is due to the property that repeating a Gaussian filtering is equivalent to filtering once with a Gaussian whose sigma is the Euclidean sum of the two sigmas), then, the difference images are determined. In the equivalent NL approach, simply the convolution between the image and the NL filters are computed. After this part, the algorithm looks for local extrema in local windows of size 3x3 or 5x5 pixels, across 3 or 5 images of the scale space sequence. The factor k is defined so that $k^m = 2$, with m integer [Lowe04]. We chose $m = 5$.

An interesting fact emerged from these tests on toy images (see also Figure 2):

1. A local maximum is found in the centre of circles of radius r , and its sigma

$$\text{is } \sigma = \frac{r}{\sqrt{2}}.$$

2. Local maxima also are found in the centre of squares and quasi-square rectangles, with a sigma equal to the minimum of the two sides, halved by two (equal to the sigma of the square inscribed in the circle of point 1).
3. Rectangles and Bars show the most irregular behaviour: local extrema are found close to the short side of rectangles, along the diagonals of the squares inscribed in them with side length equal to the short side of the rectangle or less. These *virtual* squares have one, or at most two, angles in common with the processed rectangle, and their positions correspond to different resonance scales. Small

virtual squares resonances are also found in the corners of real squares and also along the border of more irregular objects (*partial border resonance*), providing the correct sigma is employed in filtering.

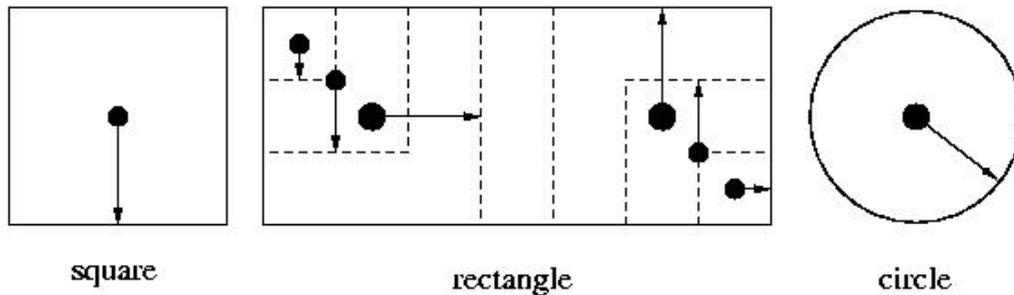


Figure 2: explication of i) local maximum found in the centre of a circle (right), ii) local maximum found in a square (left), and iii) some of the multiple local maxima found in a rectangle along the diagonals of the inscribed squares with side equal at most to the short side of the rectangle (middle).

Point 1) is very interesting for nodule detection: it means that a circular bright object of a certain radius r could be detected with a proper sigma, a sigma such that $\sigma = \frac{r}{\sqrt{2}}$ (*resonance sigma*). More in general, it means that not only a constant relative ratio between objects of different sizes and their sigma values exist, but also that an absolute relationship between the size of an object and the sigma resonance value can be found, at least for simple objects. Point 3) will become important in the discussion about the FPs detected by the system in chapter 7.

These observations formed the basis for the solution to the problem of finding a proper and easy way to estimate the size of a nodule, without employing the SRG algorithm, as we shall soon see.

Subsequently we tested the SSA on some CT images with nodules and falses, and we found out that:

1. circular and quasi-circular objects are detected by the SSA: the local maximum is positioned in the centre of the object, it is usually alone, and corresponds to a single resonance sigma;
2. less regular objects are still detected, but more than one maximum is found within a certain tolerance from the object centre: each maximum usually corresponds to a portion of the object where a circle could be individuated, and can have its own resonance sigma value;

3. relative brightness of the signal with respect to the background is not of fundamental importance: same results can be found in different conditions;
4. the radii of the detected signals, as measured on the CT images, are well approximated by the formula $\sigma = \frac{r}{\sqrt{2}}$;
5. a nodule is missed only if its resonance sigma is out of the filter range;
6. if the whole image is processed the number of detected false signals is much larger than that of FR (250 instead of 10÷50), but their characteristics are different, because only almost circular objects are found: indeed, the SSA has proved to be much more specific to circular shapes than the FR, but also less specific to bright objects, as said above (see point 3 of this list), which gives most of its falses;

After these findings, it was decided to test the SSA and the FR filters in *logical AND* combination for nodule detection. This proved to be a good choice, but, before looking at the results, another question must be addressed.

5.9 Second problem on path and a promising solution

The other question is: *is it possible to find a way to transform the image in order to further enhance the nodule-like signals? To find a transformation with higher specificity to nodules?* Indeed, there are many nodules that show low GL values, and which are therefore less enhanced, by the FR filter, than larger but less circular noise spots with higher GL values: would it be possible to enhance the former more than the latter? We know that every GL transformation is not targeted to object shape, hence is useless, because similar GL values are transformed in the same manner. A good choice would be a shape dependent transform. Well, we already know a transform with this property. It is the FR transform! Actually, a FR filtering of a FR filtered image (FR²), as the next paragraph will show, has proved to be very effective for nodule-like signal detection. It is a non linear transformation of the image which takes into account the local circularity: it can really happen that circular objects are enhanced more than brighter non-circular ones. Results are given below.

5.10 Putting all together

In Table 2 are compared results of FR and FR² on the database DB5mm, before and after the logical AND with the SSA filter.

<i>filter</i>	<i>FR 0.1</i>	<i>FR 0.16</i>	<i>FR 0.19</i>	<i>FR2 0.47</i>	<i>FR2 0.65</i>	<i>FR2 0.79</i>
<i>Nodule hit</i>	34/34	34/34	33/34	34/34	34/34	34/34
<i>Signal hit</i>	92	91	89	92	91	90
<i>FP/slice</i>	85,4	51,2	38,4	53,2	37,4	29,3
<i>FP/slice after</i>						
<i>SSA</i>	67	43,3	33,3	45,3	32,8	26,2
<i>Decrease</i>	22,00%	15,00%	13,00%	15,00%	12,00%	11,00%

Table 2: comparison of the detection results between FR and FR², before and after the logical AND with the Scale Space filter. The resonance sigma range corresponded to 2.4÷20.0 mm nodule diameter range.

The three more significant threshold values for both FR and FR² are shown. Clearly, FR² is better than FR: with the same number of 3D and 2D nodule signal hit, the average number of False Positive/slice is much lower, both before and after the SSA filtering (from 25% to 40% before, and from 22% to 33% after, depending on the threshold value). It must be noted that these variations are highly significant because the number of false signals is huge and, even if the number of 3D and 2D nodule signals is much smaller, nonetheless the comparison has been performed between the highest threshold values giving the same results. The number of 2D nodule signals is a monotone function, decreasing with increasing threshold, and are the general behaviours of these curves that have been compared. *This result shows that the first FR really acts as a nonlinear and shape-dependent GL transformation, which can improve the detection rate of the FR filter.* Because FR² is never worse than FR, we can always safely use it.

If we now analyze the effect of the logical AND between FR-based filtering and SSA filtering, we notice that there is a reduction in the number of detected FP of approximately 10 – 20%, depending on the absolute threshold value: the effect of SSA decreases with increasing FR threshold. The resonance sigmas in the searched range corresponded to nodule diameters from 2.4 mm to 20.0 mm. The lower limit is slightly smaller than the suggested value 4 mm (see chapter 1, paragraph 4) for two reasons: the first is that there are many nodules in the database smaller than 4 mm, and the second is that nodules might be composed of sub-signals, each of them resonating with its own

smaller sigma. Besides, it must be noted that, even if the maximum nodule diameter in the database is about 10 mm, the maximum searched diameter was 20 mm, in order to have a good estimate of the number of falses detected by the algorithm in the usual detection range. The comparison FR, FR² is synthesized in Figure 3.

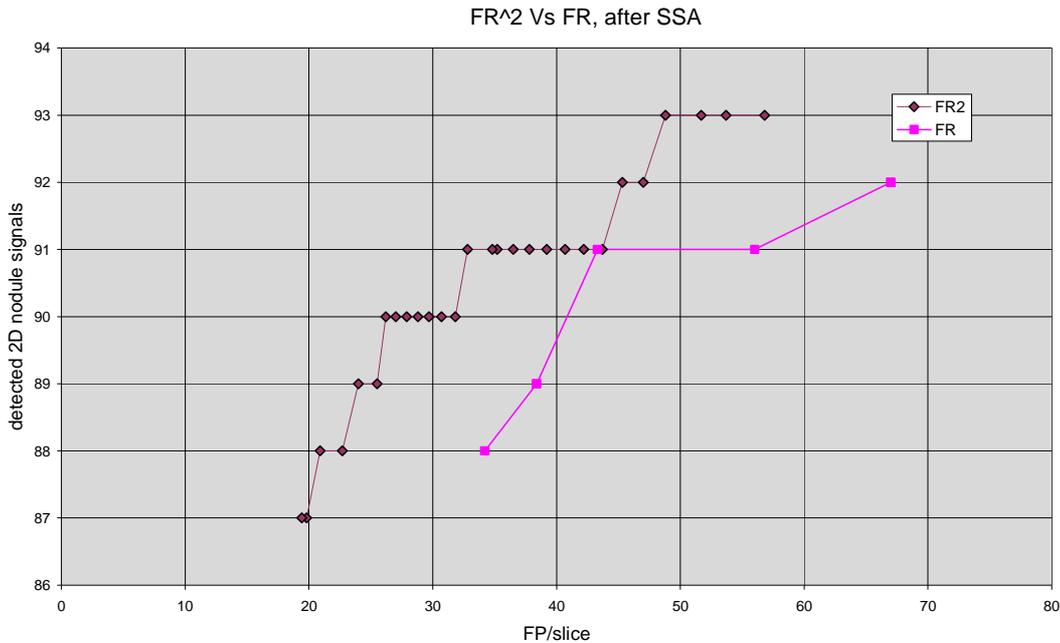


Figure 3: plot of absolute number of detected 2D nodule signals for FR and FR² Vs FP/slice, after logical AND with SSA. Threshold values go from 0.35 to 1.01 for FR², and from 0.1 to 0.22 for FR. Searched nodule range is 2.4÷20.0 mm. The superiority of FR² is evident.

5.11 Brief summary of Scale Space Approach useful properties

It is now the time to summarize the many good properties of SSA, in order to determine its impact on the detection algorithm:

1. SSA can decrease the number of false signals detected by FR or FR² of a considerable amount, depending on initial falses number and detection threshold;
2. SSA can give an estimate of the size of the detected nodule, through the relationship $\sigma = \frac{r}{\sqrt{2}}$;
3. SSA is very effective in controlling the searched nodule diameter, whose bounds can be exactly set;
4. SSA can effectively locate the nodule, because its centre corresponds to the position of the local maximum in the scale space.

Points 2 and 4 give an answer to the question arisen in paragraph 5 of this chapter, and together with point 3, allow us to get rid of the SRG algorithm, because there is no more

need to calculate the borders of the signals. Point 1 is a very welcome surprise, which strengthens the power of the Scale Space Approach.

5.12 Conclusions

The work described in this chapter has provided us with a new method for detecting 2D nodule signals in CT slices. This method, based on the Fast Radial transform used twice and on the Scale Space approach, has proved i) to have higher sensitivity and specificity, with respect to the Matching Filter method, and ii) to have many interesting properties, which will become important in the next future. As we have already seen, the SSA gives us the opportunity to abandon the Seeded Region Growing algorithm for the estimation of nodule sizes and borders, which is a critical point of the FPR part of the CAD system. Further validation of the method will be necessary in future, employing larger databases.

Chapter 6

A new CAD, part II: matching signals through slices; coarse False Positive Reduction (FPR-1)

Nodule-like 2D signals found in CT images need to be matched together to characterize 3D signals to their full extent. This operation, matching signals through slices, can also help to delete a great amount of false signals, which is a fundamental step of any CAD systems, because most of the initially detected signals do not correspond to true nodules.

6.1 Grouping 2D signals into 3D objects

Each two dimensional signal detected by previous steps of the CAD system is stored as a collection of features (spatial coordinates of signal, number and type of resonance sigmas), together with some extra-information concerning its “identity” (patient and slice number) and, only during the validation procedure, label (positive if signal corresponds to a nodule, negative if not). Since more than a single resonance sigma can be found within true nodule position tolerance, with more than a single value (see Figure 1 below), two kinds of spatial coordinates, which are given by resonance sigma positions, are considered: normal average and sigma-weighted average. For the same reason, maximum, minimum, mean and variance values of resonance sigmas for each nodule-like signal are collected. It was decided not to memorize all the single sigma values because both it would have requested too much memory and, most important, this great amount of information needs to be synthesized to be subsequently utilized.

After setting a matching tolerance (7 pixels for the present database, DB5mm), which need not to be the same as that already considered to determine the objects label (see chapters 4 and 5), 2D signals are matched across slices by simply comparing their coordinates. Beginning from the first signal in the first slice, each signal is linked with each signal in the next slice, provided their spatial positions are within the chosen tolerance.

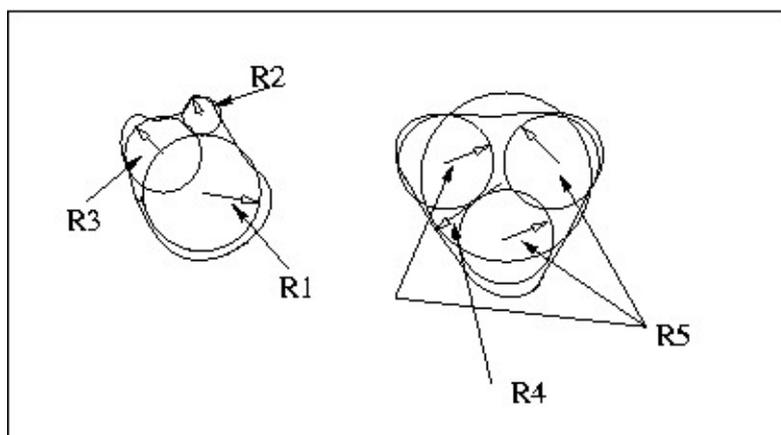


Figure 1: examples of nodules with irregular shapes and correspondent multiple resonance radii determined by sigma-resonance. Left: 3 different centres given by 3 different resonance sigmas within the same object; right: 4 centres given by 2 resonance radii within the same object.

The spatial position each signal has to be matched with is that of the last added 2D signal, not the one of the signal at the top of the chain: indeed, whilst nodules are generally found in the same position in successive slices because they are sphere-shaped, vessels, which form a large subsample of falses, moves considerably across slices. Three-dimensional linking is stopped when no correspondence is found in the last analyzed slice. It is also permitted to link signals which stay in not successive slices, to manage those cases in which there are *gaps* in the detected nodule sequence. It must be noted that object labels are not taken into account during 2D signals linking, only relative positions.

6.2 Results of 3D grouping on database DB5mm

At the end of the linking part of the algorithm, we have groups of 2D signals corresponding to objects which develop across slices in the CT scan of the patients. In the table below (Table 1) are shown some results of this 3D linking process on the database DB5mm, at different detection thresholds FRTHR (average sigma-weighted coordinates of signals has been considered).

<i>FRTHR</i>	<i>Gap</i>	<i>3D Pos.</i>	<i>3D Neg.</i>	<i>2D Pos.</i>	<i>2D Neg.</i>	<i>Positive singletons</i>	<i>Negative singletons</i>
0.51	1	34/34	33019	91	75476	2	18756
0.51	0	34/34	38866	91	75476	2	24285
0.65	1	34/34	27164	91	58773	2	16364
0.65	0	34/34	31048	91	58773	2	20264
0.79	1	34/34	22181	90	46955	3	13957
0.79	0	34/34	24757	90	46955	3	16693

Table 1: results of 3D-linking process on database DB5mm (34 nodules, 17 patients, 105 slices per patient on average), at different FRTHR thresholds. Gap 0 corresponds to no gaps in the slice sequence, Gap 1 to 1 slice gap. Singletons are signals not linked with any other signal. It is easy to see that the percentage of negative singleton is very large, whilst that of positive singletons is very small.

It is clearly seen that, irrespective to the detecting threshold FRTHR, singleton signals, that is, signals not matched to others, constitute the largest part of falses, and only a very small part of positives. Indeed, only 2 positive singletons have been found in a group of 34 nodules (about 6%), whilst a percentage of singletons varying approximately from 55 to 65% is present in the ensemble of negative 3D signals. This will be very important for the False Positive Reduction (FPR) step. On the other side, there are an average number of almost 3 2D signals per nodule, and 5 or 6 2D signals per false positive. Another important observation is that Gap 1 can reduce the absolute number of 3D falses found by Gap 0 of a percentage varying from 10 to 20%. At the same time, anyway, the percentage of singletons is reduced by the same or a greater amount, which means that the FPR impact of singletons is lowered, hence it is not clear, at this stage of development, which value is better between Gap 1 and 0. It is interesting to note that no hybrid 3D objects, that is, objects composed of negative and positive 2D signals, resulted at the end of the grouping process.

6.3 3D objects features

To further characterize the 2D signal groups, we calculated some 3D features. These features are: *length* (in number of slices), *inclination* with respect to verticality (axis parallel to CT reconstruction increment motion) of a linear fitting through centres of 2D signals, and *sigma-volume* (sum of the mean values of sigmas of each 2D signal

of the 3D object, normalized to pixels of 0.7 mm). Figures 2 and 3 below show some combinations of these features.

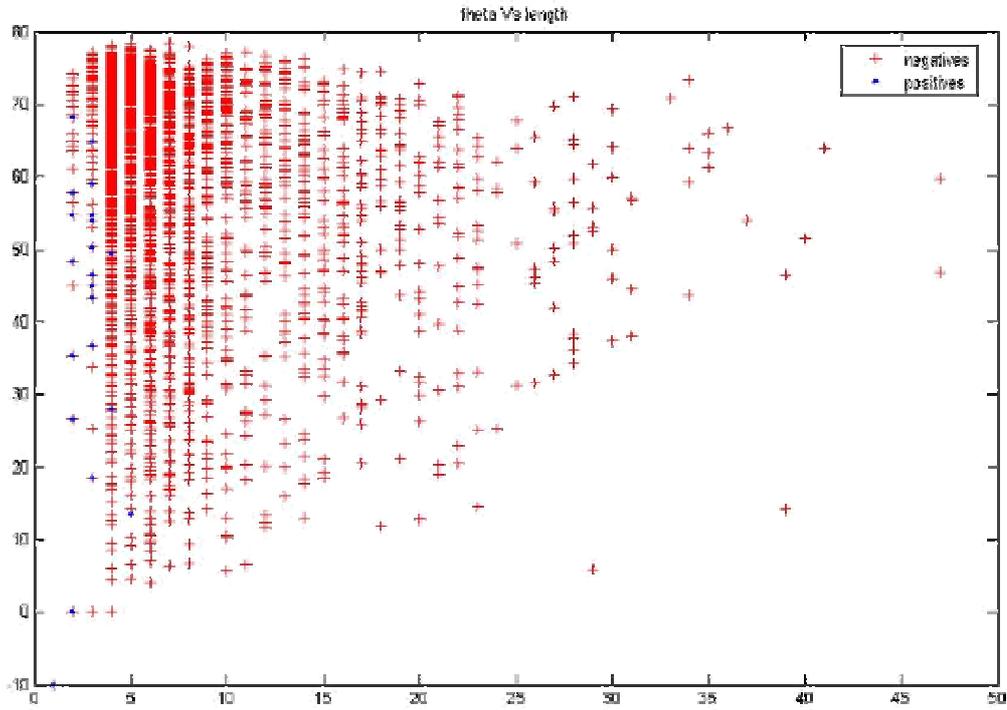


Figure 2: inclination (theta) in degrees Vs length of nodules (positives, blue), and not-nodules (negatives, red).

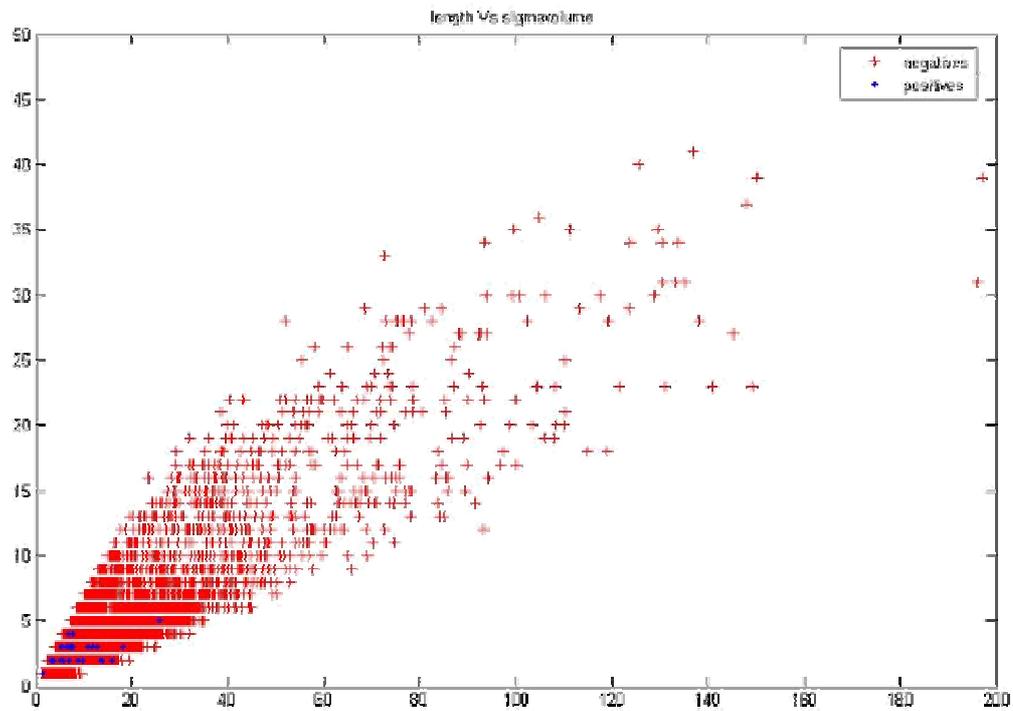


Figure 3: length Vs sigma-volume of nodules (positives, blue), and not-nodules (negatives, red).

From the above figures we can see that:

1. maximum length of nodules is 5 slices, of falses is 47 slices;
2. most of the falses are longer than 3 nodules;
3. maximum inclination of nodules is about 68° for objects 2 slices long. We could also argue that average inclination gets smaller with increasing length of nodules: indeed, nodules, being almost spherical in shape, are expected to be much less inclined than falses like vessels. The high inclination of 2-slice nodules should be due to the high variance which affects the linear regression calculation: this variance becomes less and less important when length increases, unfortunately this database is not very large, so we cannot validate this reasonable observation;
4. all but one nodules have a sigma-volume smaller than 20.

6.4 False Positive Reduction, first part (FPR-1)

All the previous observations allow us to realize a first, coarse but very effective, False Positive Reduction (FPR-1). In fact, we can decide to cut all the signals which are too short or too long, those which are too much inclined, and those whose sigma-volume is too large. In particular, too short means the signal is a singleton, and too long is related to the maximum size of searched nodules. Because we are looking for nodules smaller than 20 mm, we cut objects longer than 8 slices (in this database the overlapping amount between slices is 2 mm, hence it is quite reasonable for a 20 mm long object to be no more than $7 \div 8$ slices in length). At this point it is important to note that the database at hand is too small to allow a Cross-Validation procedure; hence we are forced to choose all these FPR-1 parameters so that they are no stricter than what we could reasonably estimate for the general distribution of nodules we are looking for. In other words, as we cannot validate these parameters, we have to accept a larger amount of false positives than what we could really eliminate by looking at this database alone. This is the reason why we do not accept signals more inclined than 70° , when long 2 or 3 slices, 60° when long 4, 40° when long more than 4, and why we cut all signals with sigma-volume larger than 40: in both cases, smaller thresholds could have been set, but with the risk to be too much fit to this limited distribution. The general prescription is to apply lax criteria and to be not too much specific to the dataset: other researchers have adopted this approach [Gurcan02], when the limited size of databases could not allow putting up a validation procedure.

6.5 FPR-1 results on DB5mm

Results of this FPR-1 procedure are shown in the table below (Table 2).

<i>FRTHR</i>	<i>Gap</i>	<i>3D</i> <i>Pos.</i>	<i>3D</i> <i>Neg.</i>	<i>2D</i> <i>Pos.</i>	<i>2D</i> <i>Neg.</i>	<i>After FPR-1</i> <i>3D Pos.</i>	<i>After FPR-1 3D</i> <i>Neg.</i>	<i>Survived</i> <i>Neg.</i>
0.51	1	34/34	33019	91	75476	32	9003	27.3%
0.51	0	34/34	38866	91	75476	32	9793	25.2%
0.65	1	34/34	27164	91	58773	32	6611	24.3%
0.65	0	34/34	31048	91	58773	32	7027	22.6%
0.79	1	34/34	22181	90	46955	31	5105	23%
0.79	0	34/34	24757	90	46955	31	5232	21.1%

Table 2: effect of FPR-1 on signals detected with threshold values FRTHR. The amount of removed falses is approximately from 70 to 80%, at the expense of only 2 or 3 nodules over 34.

From Table 2 it is easily seen that approximately 70% to 80% of false nodules are eliminated by this FPR-1 step, and at the same time that only 2 or 3 nodules are lost, over 34, because they are singletons. These are very good figures, and the fact that the decrease in negatives is large, no matter the value of the detection threshold FRTHR, is a particularly good result; anyway, further validation is necessary to clearly understand if heavier cuts could be done. Again, impact of Gap parameter cannot be judged: values 0 and 1 seem to be quite equivalent, at this stage.

6.6 Conclusions

Matching of 2D signals across slices is a very important step of the CAD system, because it allows to process signals in their 3D extent. The 3D characterization of detected objects has proved to be particularly well-suited for the FPR step, which has received a great benefit from the 3D linking procedure, as can be seen by comparing the FPR results of the first CAD system described in Chapter 4, in which, for example, there is no impact at all of singletons. Of course, further validation on a larger and independent database will be necessary to assess the absolute performances of this False Positive Reduction step.

It must be stressed that the 3D linking procedure and the following FPR-1 step are based on the use of the Scale Space Approach described in Chapter 5. The linking procedure makes use of the signal positions as detected by the corresponding resonance sigmas of the SSA, and one of the cuts of FPR-1 is based on resonance sigmas as well. As already stated in Chapter 5, the possibility to getting rid of the SRG algorithm has had a fundamental consequence on the development of the algorithm, in terms of

simplification and reliability. What needs now to be tested is whether or not abandoning the geometric features has been a good choice.

Chapter 7

A new CAD, part III: fine False Positive Reduction (FPR-2)

As already stated, False Positive Reduction is a fundamental step of CAD systems of any kind. In this particular case, FPR is in charge of discriminating true nodules from false ones among all the signals detected by our system. In fact, the system performances depend both on the sensitivity and specificity of the initial filtering part, and on the discriminative ability of the FPR part. We have previously seen, in chapters 5 and 6, that our system is composed of an initial filtering part, deputed to detect nodule-like signals in CT slices, and of a coarse FPR part, able to delete more than 70% of the falses detected by the system. Unfortunately, the remaining amount of falses is still very large, thus another FPR step is necessary (see for example chapter 4 of this thesis and [Armato02], [Suzuki03]). This step will be based on a Support Vector Machine classifier, and will comprise two sub-steps: one is the classification of each 2D signal; the other is the final labelling of each group of 2D signals, which have been given the identity of a single 3D object through the linking process described in the previous chapter. Eventually, only the 3D objects judged as nodules will be prompted to the final CAD user.

7.1 Preparing for signals classification: choosing Regions of Interest and features

We wrote in paragraph 6 of chapter 5 that we don't need to use geometric features for classification, because there are many different kinds of features that have proven to be very effective for image classification, such as Gray Level values, wavelets,

ranklets, etc. What is necessary to do, when using this kind of features, is to find a square area around the object to be classified (Region of Interest: ROI) and use the features derived from this small image as features for the classifier. We have chosen to use as features in this work the Gray Level values of the ROI. This type of features has been used with success to classify microcalcifications and masses in CAD for mammography [El-Naqa02][Masotti05], and could also be considered as the coarse version of the Support Vector Regression Filtering approach described in [Angelini05]. The only problem would be that of precisely determining the size of the signal to be classified, in order to crop a ROI which is adequate to the object size. Indeed, the aim is to obtain a ROI not too large or too small with respect to the signal, because the former would make the signal to be “lost” in the background, the latter would focus only on the signal, without consider its relationship with the background. Balance is not easy to reach. The SSA helps us in estimating the size of the signal detected in the slice, through the relationship found in chapter 5, paragraph 8, based on the resonance sigma

value: $\sigma = \frac{r}{\sqrt{2}}$. We could then set the ROI side proportional to the radius r , by

introducing the formula, depending on the parameter *extraside*:
ROI side = $2 * (1 + \text{extraside}) * 2r$, (see Figure 1).

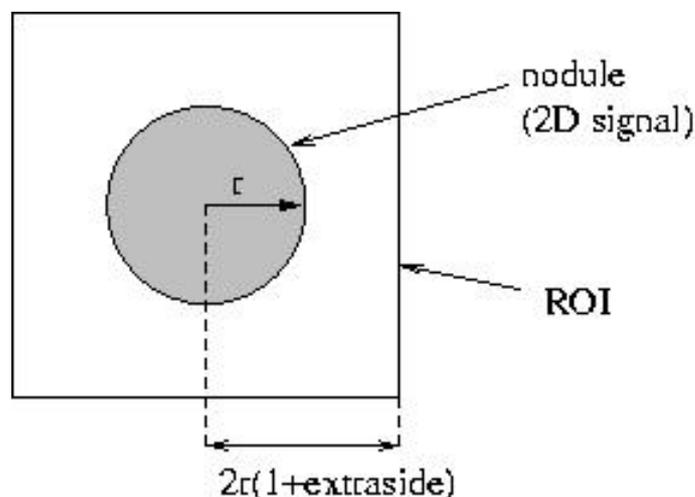


Figure 1: relationship between ROI side and nodule radius.

By taking into account previous experiences regarding the classification of ROIs containing mass lesions [Angelini05], we already knew that the area of a lesion with complex texture, as a mass is, needs to be not smaller than 17÷22% of the total area of the ROI. Since a nodule has a considerably less complex texture than a mass, we made the hypothesis that the background plays a considerably more important role for nodule

classification than what it does for masses: this is the reason why we decided to start with the value $extraside = 0.05$, which makes the signal area be about 19% of the ROI area, and then compare it with the values 0.15 (18.2%), 0.24 (17.5%).

Afterwards, all ROIs are resized by means of a bilinear algorithm to a common ROI size (parameter *RESIZE*, indicating the ROI side length, in pixels), and then their GL pixel values are put in form of a vector, by simply building a single column vector made by the columns of the ROI matrix. These vectors are the feature vectors for the SVM classifier. Classification of each ROI is the first step of FPR-2: after that, each group of 2D signals, corresponding to a 3D nodule-like signal initially detected by the system, is given a final label, positive if it is judged a nodule, negative if not. The *RESIZE* parameter was chosen among the values 11, 15, 19, 23, respectively corresponding to vectors of length 121, 225, 361, 529. We must note in fact that it is important not to reduce too much the side length of the ROIs, because valuable information could be lost during the process, so that the resized ROI would be useless for classification purposes. On the other side, we must remember that the risk referred to as *the curse of dimensionality* is always present in learning algorithms, when the number of training vectors is close or smaller than the classification space dimensionality [Haykin99], even if we know that SVM is much less prone to this risk than all other learning algorithms [Vapnik95][Campanini04]. In conclusion, because the smallest ROI size detected by the algorithm was 11 pixels, and because the number of positive 2D training vectors is less than 100, we decided to choose the above values for the parameter *RESIZE*.

7.2 Classifying 2D signals: 7-Fold Cross Validation and 90° rotations of ROIs

Consider now the classification of 2D signals. Because the database DB5mm is a small one, a Cross-Validation procedure was considered for training and for performance estimation. The set of survived 3D signals, for $FRTHR = 0.65$, is composed of 32 nodules and approximately 7,000 falses, corresponding to 89 2D positive signals and 16,999 (Gap 1) or 17,394 (Gap 0) negative ones. This $FRTHR$ value was chosen because it seemed a good compromise between nodules and falses numbers: as already noticed, a larger database will be necessary to choose the optimal value of parameters. Unfortunately, of the 13 patients with nodules, 2 contain 17 nodules of the total number of 32, hence a Leave-One-Out procedure was not possible, and 7-Fold CV was chosen instead: this is a good trade-off between the necessity to

have balanced sub-groups in the CV, that is, with similar numbers of positive and negative signals, and the prescription of CV experience, saying that the number of sub-groups should be closer to 10:20 [Kohavi95]. The 4 patients without nodules were used to make the number of signals as balanced as possible across the sub-groups. Before proceeding with the CV, all the data vectors underwent a whitening procedure, “pixelwise”: the distribution of each pixel, being a dimension of the feature space, was individually normalized to zero mean and unit standard deviation (see also chapter 4, paragraph 9). We tried a polynomial kernel for SVM, with low degree (2, 3), that is known to always guarantee good performances with low overfitting risk, at a low computational cost [Campanini04].

Further, to increase the number of positive signals and, at the same time, to introduce some *rotational invariance* in the classifier generalization capability, all positive ROIs in each training procedure were rotated 3 times by 90° and the new correspondent vectors added to the training. With this operation, the number of positives in training is multiplied by 4 (from 89 to 356). Moreover, this procedure was also introduced into ROIs classification, by considering the 4 labels of the 4 90° rotated views of each ROI. Different thresholds can be set for this labelling procedure: 2, 3 or 4 positive labels can be required to give the final positive label to the ROI.

7.3 Labelling 3D objects

Consider now the labelling of 3D objects. Their labels must depend on the individual labels of each of their 2D signals. If we look at these 2D signals, we notice that not all have the same *quality*, that is, some of them really resemble a nodule, while others do not, especially those at the beginning and at the end of the 3D object (see Figure 2 below).

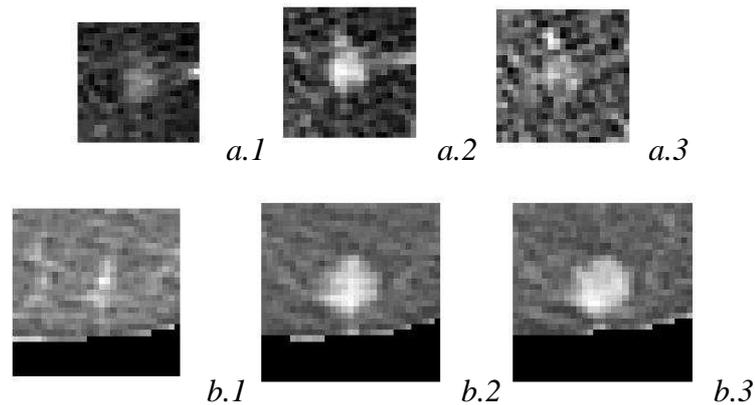


Figure 2: examples of two 2D nodules sequences relative to two different 3D nodules. Differences in quality are evident: the last signal of sequence *a* is much noisier than the first two ones; the first signal of sequence *b* is very small and difficult to spot.

This can be easily explained: the first and last 2D signals usually come from slices which have cut the nodule in its top or bottom, so that in the slice it has been reconstructed together with some of the lung tissue positioned over its top or under its bottom. This makes the nodule be very difficult to see, and smaller in size than its maximum extension. To try to overcome the fact that top and bottom low quality 2D nodules present a high risk to be classified as negative signals, we decided to use not strict procedure when labelling a 3D object: it is sufficient that only a fraction of its 2D signals are classified as positive to give it a positive label. This fraction must depend on the 3D object length. This is a sort of Majority Voting (MV) algorithm [Kittler98].

7.4 7-Fold CV results – part I

In this section we shall see the first final results of the CAD system. The procedure of CV on the dataset of 356 2D positives and 16,999 2D negatives was performed to optimize the following parameters in their respective ranges: *extraside* = 0.05, 0.15, 0.24; *RESIZE* = 11, 15, 19, 23 pixels; *Gap* = 1, 0 slices; SVM kernel: *polynomial*, with degree 2, 3. Because we found that all the training processes terminated with complete separability of the two classes, both the SVM *cost parameter* C and the SVM *unbalancing parameter* $C^{\pm} = (\text{number of negatives})/(\text{number of positives})$ were useless, hence were no more considered.

The MV algorithm was used with these settings: a 3D object is given a positive label if at least one signal is positive, in case it is long up to 4 signals, and if at least two signals are positives, in case it is longer than 4 slices (remember that the maximum allowed signal length in the database DB5mm is 8 slices). From the histograms in

chapter 6, we know that the 3D positive objects in this database are all long less than 5 slices, except one. The average length is 2.8 slices for positives and 3 for negatives objects, even if the maximum length of negatives is 48 slices. It is clear that, for this database, the behaviour of the algorithm on objects no longer than 4 slices is what really matters for what concerns performances on true nodules; on the other side, as already stated in the previous chapter, we tried to obtain an estimate of the FP rate of the CAD system for nodules in the range 3÷20 mm: this is the reason why we tried to use lax criteria in the FPR-1 phase and in this labelling phase as well, by not eliminating all signals longer than 5 slices. We initially compared the above MV settings with a strict labelling procedure, requiring all 2D signals to be positive to give the 3D object a positive label, but, as the weak labelling results below will show, this was really too strong a requirement.

In the table below (Table 1) are shown the best results of the 7-Fold CV procedure.

<i>Gap</i>	2D pos.	2D neg.	3D pos. average	3D pos overall average	3D neg	FP/Pat. - average	SVs number: positives	SVs number: negatives
1	33-37 /89	350-390 /16999	23/32	23/34	320-350 /6611	19.4	~260/330	~750/13500
	0.40	0.022	0.719	0.676	0.05		75÷80%	5.5%
0	35-37 /89	350-400 /17394	23.7/32	23.7/34	320-360 /7027	19.9	idem	idem
	0.40	0.022	0.74	0.699	0.05			

Table 1: results of the 7-Fold CV (*extraside* = 0.05, polynomial kernel with *degree* = 2), averaged over *RESIZE* parameter values, for both *Gap* 1 and *Gap* 0. The *overall* column takes into account the total number of nodules of the database. The second and the fourth rows report the fractional number of positives and negatives. In the last two columns are shown the average number of Support Vectors per class, with respect to the total number of training vectors in the 6 sub-groups (the other one is used as test).

These results have been obtained with parameters: *extraside* = 0.05; *kernel* = polynomial, with *degree* 2; number of positive labels over 4 necessary to give a positive label to a 2D signal (ROI): 2; MV parameters: at least 1 positive for objects long up to 4 slices, at least 2 for longer objects. Moreover, these results have been averaged over the *RESIZE* parameter values, because no clear indication was obtained for an optimal resizing value among the proposed four ones.

The other values of *extraside* gave no better results, sometimes worse: approximately equal or slightly lower 2D positive rates with about 22÷25 FP/Patient.

Polynomial kernel of *degree* 3 gave approximately the same positive rates with 150% FP/Patient.

As can be seen from Table 1, no differences can be found in the performances of *Gap 1* and *Gap 0*.

It is important to note that the performances of the classifier over 2D signals are not brilliant from an absolute point of view, because only 40% of positives are hit, but are quite good from a relative point of view, because only 2% of negatives receive a wrong label. This means at the same time that: i) the GL features and the ROIs database carry an amount of information useful to distinguish the two classes; ii) this amount of information might not be sufficient to improve the performances. At this point, it is not possible to know whether the database or the features need first to be improved. From previous experiences, we could argue that it is the lack of data that mostly affects the performances.

In conclusion, this system configuration can find almost 70% of nodules in the database range 3÷20 mm at a cost of about 20 FP/Patient. To compare these figures with those obtained by the system described in chapter 4, whose results were 70% of nodules with 2.1 FP/slice, we must consider that the total number of 2D falses of the present system is approximately 350÷400, and that they have been found in 17 patients with 105 slices on average each: this gives 0.2 FP/slice. The present system, according to the 7-Fold CV results, is ten times better than the systems of chapter 4.

7.5 7-Fold CV results – part II: choosing positives for training

It is now important to consider again what was noted in paragraph 7.3: not all the 2D nodule ROIs have the same *quality*. Top and bottom ROIs present a lower quality, when compared to *inner* ROIs. Because CV is a procedure which considers all the samples in the database alternatively for training and testing, this fact is very important and leads to the observation that a better training dataset might be prepared, by choosing only the high quality samples. In other words, the multiple learning procedures of the CV should be performed with a smaller positive set, in which all the ROIs must have been chosen according to their qualitative aspect. This was done, and 54 ROIs were chosen among the initial 89 (resulting in a set of 216 2D positives): about 40% of the detected positive signals were judged low-quality ones and not considered for training (but were for test). Results are given in Table 2.

Gap	2D pos.	2D neg.	3D pos. average	3D pos overall average	3D neg	FP/Pat.-average	SVs number: positives	SVs number: negatives
1	28-31/ 89	170-200 /16999	21.3/32	21.3/34	150-180 /6611	9.6	~150/185	500/13500
	0.32	0.011	0.666	0.625	0.025		80%	4%
0	24-31 /89	200 /17394	20.8/32	20.8/34	170 /7027	10.1	idem	idem
	0.32	0.011	0.65	0.610	0.024			

Table 2: results of 7-Fold CV (*extraside* = 0.0.5, polynomial kernel with *degree* = 2) with chosen positives in training, averaged over *RESIZE* parameter values, for both *Gap* 1 and *Gap* 0. The *overall* column takes into account the total number of nodules of the database. The second and the fourth rows report the fractional number of positives and negatives. In the last two columns are shown the average number of Support Vectors per class, with respect to the total number of training vectors in 6 sub-groups (the other one is used as test). Clearly, results are better than those reported in Table 1.

Clearly, the system with chosen positives in training reaches better performances: indeed, even if the absolute hit rate is lower, the FP/Patient figure is halved! Moreover, the total number of Support Vectors (SVs) is decreased to approximately 60÷70% of previous level (Table 1), even if the absolute number of samples in the positive class is now 60% of the initial one. This means that [Vapnik95] the two classes are now more separable, or separable with less effort, which in turn means that the positive class is now more dissimilar to the negative class. The fact that without the selection of positives there were 2 times FP/Patient and only a small increase in correctly classified positives, signifies that in the 89 positives database there are many signals very close to the other class: the decision to discard them was hence right. Besides, this fact shows that the information carried in the training database is of fundamental importance to obtain good performances, and supports our observation of paragraph 7.4, that enlarging the database would be more important, at present, than trying different classification features, whose results could, also, not easily compared on this small database.

7.6 Introducing *Multirotation* in training

As a consequence of the previous result, we tried to find the way to use at best the information of the database. The natural way to do it was to force the procedure considered in paragraph 7.2 for introducing rotational invariance and increasing 4 times the number of positive ROIs in training. This time, we decided to add to the training set more rotated versions of each ROI, before applying the usual 90° rotation procedure. Each ROI is rotated of 15°, 30°, 45°, 60° and 75°, before being cut from the original image (otherwise there would have been problems with the image corners). The procedure (*Multirotation*) gives 24 times the number of original positives to be used in

training ($54 \times 6 \times 4 = 1,296$). The new Multirotation training procedure can also affect the classification procedure of 2D signals. Two options are hence possible: going on with the usual 4-rotation labelling procedure, or try to use a Multirotation based one. We tried both, beginning with the usual 4-rotation procedure.

7.7 7-Fold CV results – part III: effects of Multirotation in training

In the table below (Table 3) are reported the results of the 7-Fold CV with chosen positives, Multirotation procedure in training and usual 4-rotation procedure in test. In the latter, two different thresholds were considered to give a positive label: at least 2 over 4, or at least 3 over 4. Only Gap 1 was considered, because results are similar for both values 0 and 1.

With Multirotation								
4-rotation threshold	2D pos.	2D neg.	3D pos. average	3D pos. overall average	3D neg.	FP/Pat. average	SVs number: positives	SVs number: negatives
At least 2	45-47 /89	380-410 /16999	23.7/32	23.7/34	320-345 /6611	19.8	~380/1150	650/13500
	0.52	0.022	0.74	0.697	0.04		33%	5%
At least 3	36-40 /89	180-200 /16999	22/32	22/34	155-165 /6611	9.5	idem	idem
	0.45	0.011	0.68	0.65	0.04			
Without Multirotation								
At least 2	28-31 /89	170-200 /16999	21.3/32	21.3/34	150-180 /6611	9.6	~150/185	500/13500
	0.32	0.011	0.666	0.625	0.025		80%	4%

Table 3: comparisons between 7-Fold CV with and without Multirotation in training, both employing usual 4-rotation labelling in test, with “at least 2” and “at least 3” thresholds. The introduction of Multirotation has increased the 2D performance rate but also the 2D FP rate, in the case of “at least 2” labelling threshold. The labelling threshold “at least 3” gives the best results.

From the table above it is easy to see that the Multirotation training is superior to the previous 4-rotation training: indeed, it is superior from the point of view of the number of 2D positive signals correctly classified, and at the same time it finds the same number of FPs as before, when the labelling threshold is in the configuration “at least 3”. This fact means that the classifier has learned better than before the characteristics of the two classes, positives and negatives, due to the gain in information given by the Multirotation procedure, with which the nodule signals are “seen” by the classifier during the training step from many orientations. The absolute numbers of SVs have both increased: the positives almost 3 times, the negatives about 30%. This fact indicates that a better positive class description has forced the learning algorithm to ask for more information about the negative class. However, the fact that the two classes are easily separated in the training phase, but absolutely not separated in the test one,

together with the high difference between the positive and negative class percentage figures and the still larger absolute number of negative SVs with respect to the positive ones, denote that more information should be added to reach a better separability. In particular, the nodule class is certainly underrepresented in the training set.

7.8 Introducing Multirotation in test phase

The next step of the system development was the introduction of the Multirotation procedure in the 2D labelling phase. 24 views of the same ROI are classified by the SVM algorithm to obtain the most probable label between “nodule” and “not-nodule”. In order to optimize the threshold of this parameter, we first of all considered the following histograms (Figures 3 and 4): they show how many 2D signals has received a certain fraction of positive labels over the 24 (data shown for *RESIZE* = 11).

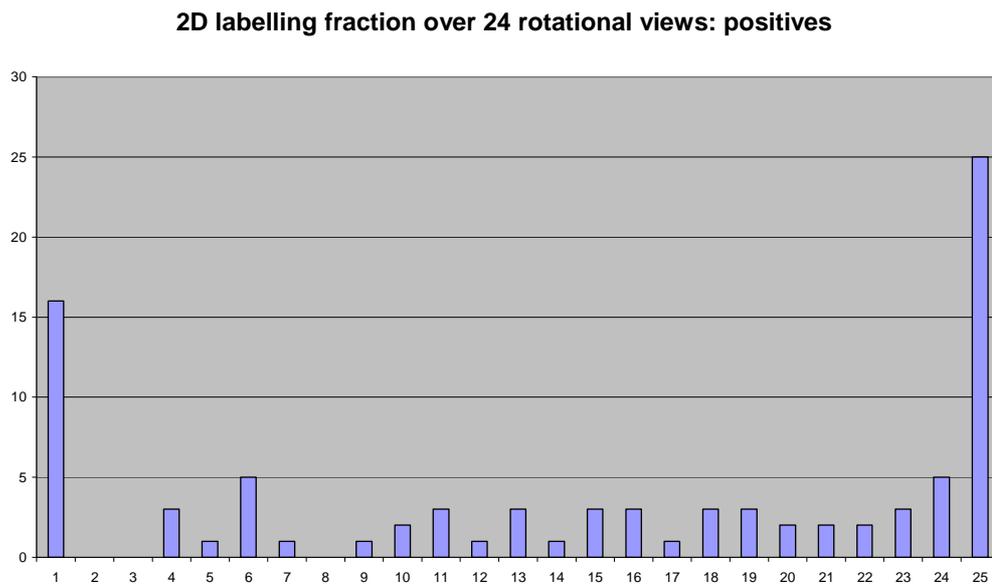


Figure 3: this histogram shows how many 2D positive signals have received a certain number of positive labels, from 0 (column 1) to 24 (column 25).

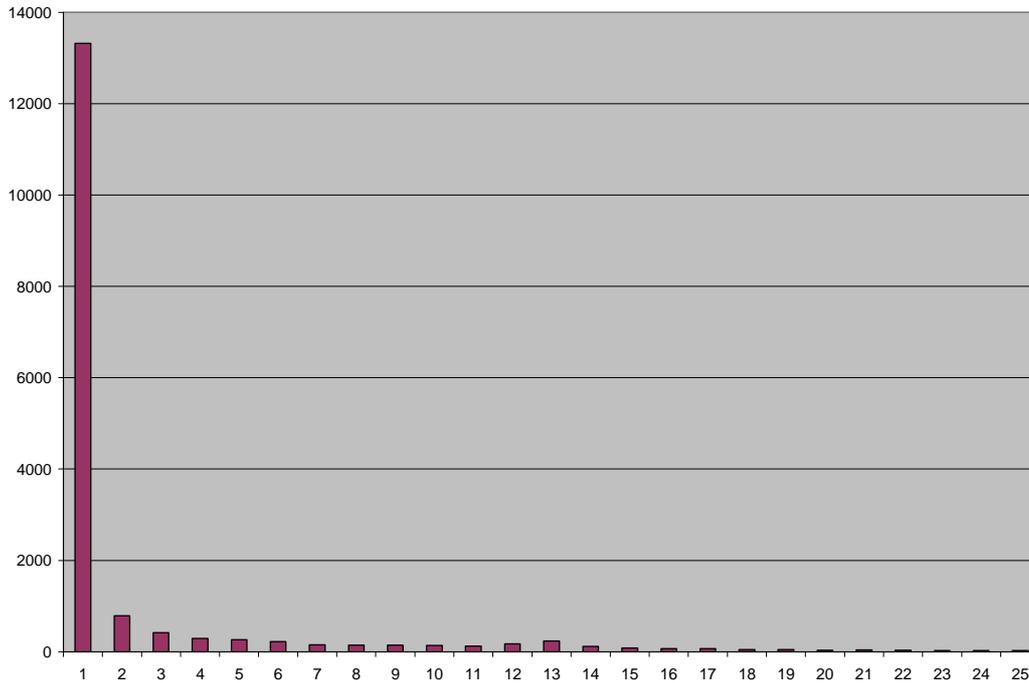
2D labelling fraction over 24 rotational views: negatives

Figure 4: this histogram shows how many 2D negative signals have received a certain number of positive labels, from 0 (column 1) to 24 (column 25).

Figure 4 shows that the classifier is able to give 0 positive labels over 24 to 13,320 negative ROIs over 16,999 (78.4%): moreover, only 593 negatives are given at least 13 positive labels. These are the false positives that will eventually contribute to the FP/Patient rate of the system. Figure 5 shows that 25 ROIs over the initial 89 are given 24 positive labels over 24 (28%), that 16 over 89 are given no positive labels at all (18%), and that 53 over 89 are given at least 13 positive labels over 24. The monotonically decreasing number of positive labels given by the classifier to the negative ROIs, together with the less regular but almost mirror behaviour of the positives histogram, tell us that we have to analyze the performances coming from all fractional values between 12 and 24 with the 7-Fold CV procedure, in order to obtain the optimized system. Results are given below.

7.9 7-Fold CV results – part IV: effects of Multirotation in test and final FROC curve

The next figures (Figures 5 and 6) show the effect of varying the labelling threshold from 12 to 24 over the positive and negative ROIs.

Correctly classified 2D positives at varying labelling threshold

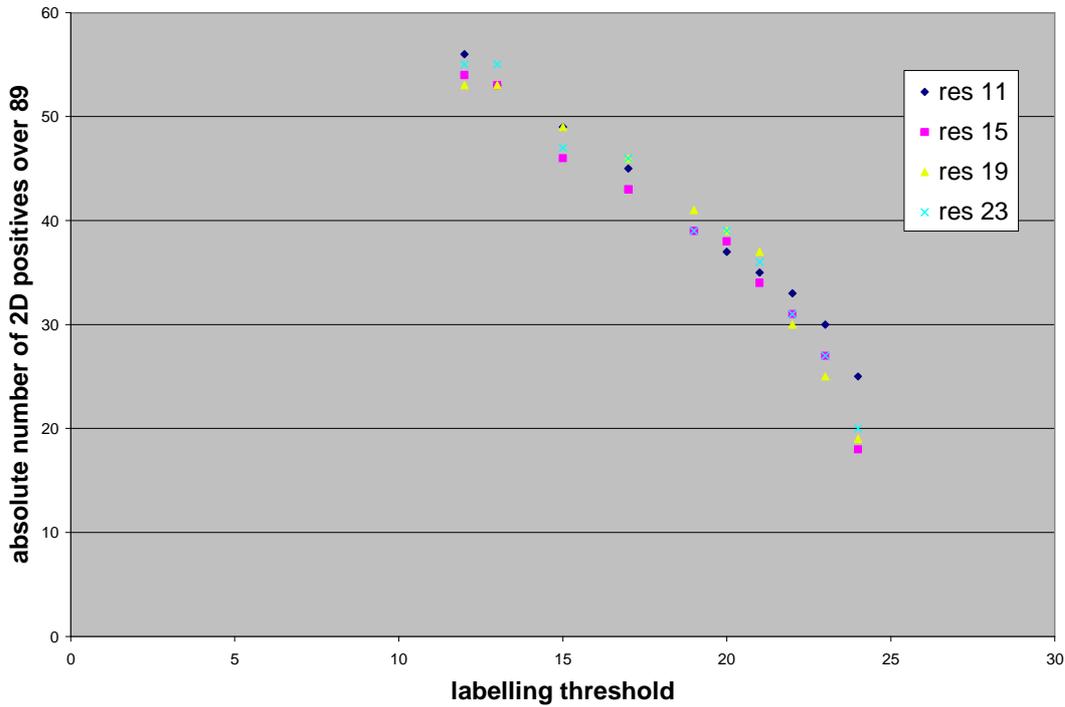


Figure 5: absolute number (over 89) of correctly classified 2D positive signals at varying labelling threshold from 12 to 24. In the legend, “res N” stays for “ROI resized to N pixel”.

Wrongly classified 2D negatives at varying labelling threshold

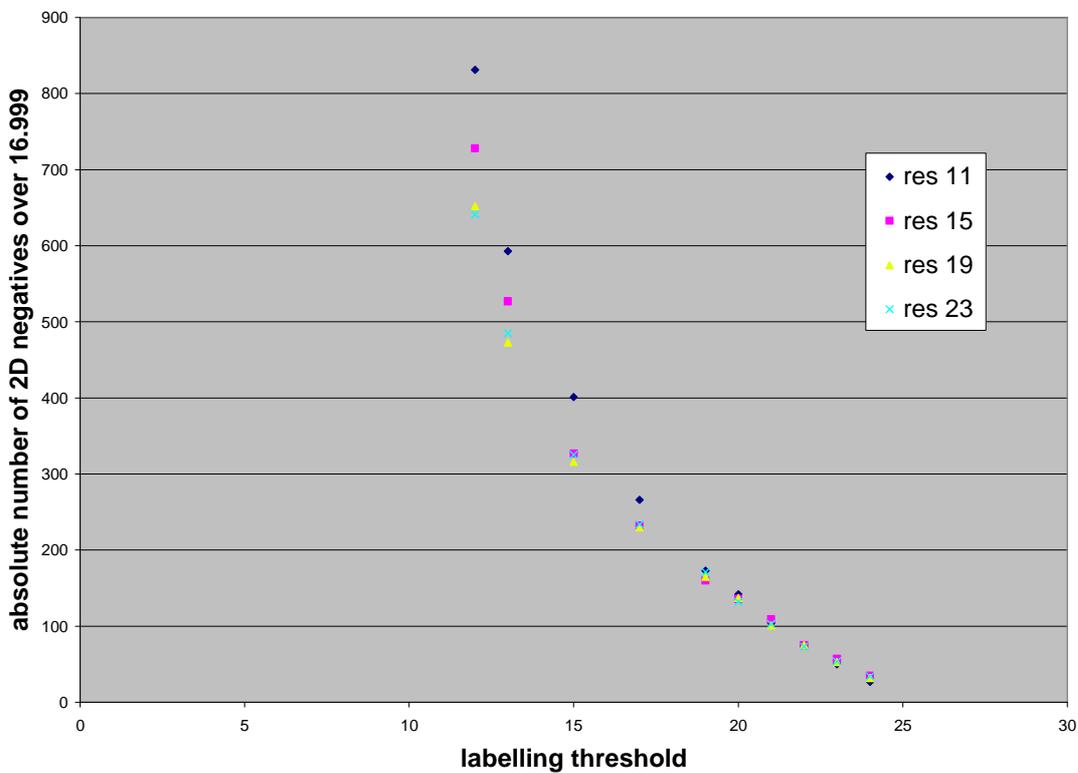


Figure 6: absolute number (over 16,999) of wrongly classified 2D negative signals at varying labelling threshold from 12 to 24. In the legend, “res N” stays for “ROI resized to N pixel”.

The system behaviour seems in general to be not particularly dependent on the resizing parameter (*RESIZE*), except for the value $RESIZE = 11$, which presents a larger number of false positives for labelling threshold values from 12 to 17, and a larger number of true positives for threshold values equal to 22, 23, 24.

As already noted, the smallness of the database makes it difficult to clearly determine some fine details of the system behaviour, hence we decided to simply calculate its average behaviour over the *RESIZE* parameter to derive an FROC (Free Response Operating Characteristics) curve [Metz86] to summarize the system performances. A FROC curve is a plot of True Positives (TP) rate, that is, fraction of detected nodules Vs the average number of FP/Patient, and is widely used in the CAD community as a tool for a quick and effective representation of systems performance. In the table below (Table 4) we can find the performances of the system averaged over the resizing parameter.

2D labelling threshold (over 24)	12	13	15	17	19	20	21	22	23	24
2D pos	54,5	53,5	47,75	45	39,5	38,25	35,5	31,25	27,25	20,5
2D neg	713	519,5	342,25	240,5	167,25	136,75	104,25	74,5	53,5	32
3D pos	27	26,75	24,75	23	22,75	22,5	21,5	20	18	16
3D pos: fraction	0,794	0,787	0,728	0,676	0,669	0,662	0,632	0,588	0,529	0,471
3D neg	572	426	291	206	147	121	91	65	46	27
2D FP/slice	0,399	0,291	0,192	0,135	0,094	0,077	0,058	0,042	0,03	0,018
3D FP/PAT	33,6	25,0	17,1	12,1	8,6	7,1	5,3	3,8	2,6	1,6

Table 4: absolute and fractional numbers of positive and negative signals detected by the system at varying labelling threshold. The database is formed by 89 2D positives, 16,999 2D negatives, 34 3D nodules and 17 patients. Results are averaged over the 4 different values of the *RESIZE* parameter. The number of 3D FP/Patient is calculated dividing the number of 3D negatives by the number of patients, whereas the number of 2D FP/slice by dividing the number of 2D falses by the number of patients times the average number of CT slices per patient (105).

The average percentage of correctly classified 2D positives varies from 23% to 61%, with an average percentage of wrongly classified 2D negatives varying between 0.2% and 4%. Correspondingly, 47% to 79% of nodules are detected, with 1.6 to 33.6 FP/Patient (or 0.02 to 0.4 FP/slice).

These averaged results allow us to plot the FROC curve of the system over 34 nodules in 17 patients of the database DB5mm (see Figure 7 below).

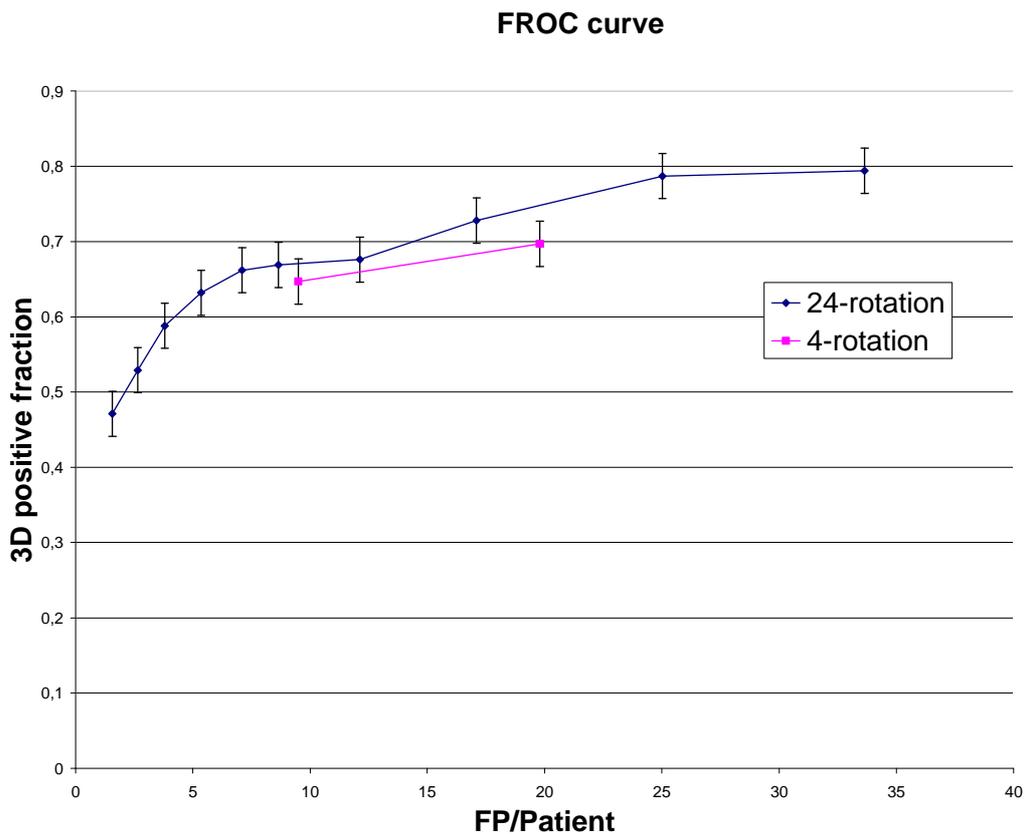


Figure 7: FROC curve of the CAD system for the DB5mm database. The 24-rotation curve refers to Multirotation algorithm used in test, with labelling threshold going from 12 to 24, whilst the 4-rotation curve refers to the usual 4-rotation algorithm, with labelling threshold equal to 2 and 3. In both Multirotation procedure was employed in training. Error bars correspond to the standard deviation: 0.03.

The FROC curve shows different working points of the system, depending on the labelling threshold. The first important result of Multirotation based test is in fact a finer tuning of the system, with respect to the 4-rotation based test. Furthermore, with the Multirotation procedure the system reaches better 2D positive rates at fixed 2D or 3D negative rates, as can be seen by comparing Tables 3 and 4. Errors bars in the FROC curve, about 0.03 in absolute value, approximately corresponding to 1 nodule, simply refers to the standard deviation of the 4 different curves given by the 4 different values of the *RESIZE* parameter. However, if we take into account that these results have been obtained from a Cross Validation procedure on a small database of only 34 nodules, we must conclude that larger errors might affect these performances, and that these errors come from bad sampling of the database with respect to the *real*, and unknown, nodule distribution. Unfortunately, these effects can't be directly estimated. As an alternative, we could consider again, as in chapter 4, the binomial proportion interval estimation

[Berger95][Brown02]: at 95% Confidence Interval, the error is approximately 3÷6 nodules (9÷18%). Keeping in mind this figure, we notice that these results are comparable with those found in CAD literature, some of which are reported in chapter 2: consider for example [Armato02], [Lee01], [Gurcan02].

7.10 False Positives analysis

By direct visual inspection, it is possible to note that the set of FP signals can be divided into three categories:

- a. ROIs found at the outer border of the lung area, representing normal lung tissue;
- b. ROIs around large and non-circular bright zones, usually part of vessels;
- c. ROIs around small circular vessels.

Whereas falses of type *c* are quite reasonable and constitute the largest subgroup, the other two types, though less common, being more dissimilar to relevant nodules, should be avoided as much as possible in order to reduce CT reader distraction to a minimum. This will be the goal of future work, anyway, it can be argued that falses of class *a* are related to SSA and FR^2 : indeed, in our initial test with SSA we noted that there can be resonance not only with whole objects (circles or squares), but with parts of them (*partial border resonance*: see chapter 5, paragraph 8). When the segmentation pre-processing leaves some very bright pixels on the border of the lung area, it can happen that they are detected by the FR^2 algorithm and, if in the same area a partial border resonance with lung tissue is found by the SSA, then the ROI will be considered for further processing. The fact that the training process is based on a small subgroup of the real and unknown nodule distribution, and that some of the used ROIs contain a very small nodule surrounded by lung tissue, is probably responsible for the detection of this kind of FPs. Similarly, even type *b* and *c* of falses will probably be reduced by the exploitation of a proper large database. The figures at the end of the chapter show examples of nodules and of FPs (*RESIZED* = 19, labelling threshold = 15).

7.11 Conclusions

In this chapter we have seen the last part of the CAD system, the FPR-2 step, based on Gray Level features and a Support Vector Machine classifier. Even if the employed database is not large enough to perform an independent test, the 7-Fold CV procedure allowed us to clearly show that:

1. the GL features have good discriminative capabilities;

2. the introduction of the Multirotation procedure in training makes the classifier extract a great amount of information from the database, greater than the previous 4-rotation procedure;
3. the introduction of the Multirotation procedure in the test phase gives the system a larger number of working points, and better results;
4. the kinds of falses made by the system can be divided into three subgroups, one made of signals very similar to nodules, the other two related to the system structure.

As already noticed, all the results need to be further validated on larger databases, but it must be remembered that, even if the nodules in the database DB5mm ranged from 3 to 10 mm, we searched nodules in the range 3÷20 mm, in order to obtain a good estimation of the FP/Patient figure of the system. Moreover, the parameters in the FPR-1 phase were chosen to be *lax*: not tailored to the positive distribution in the database, but reasonable for the “real” positive distribution. As a consequence, overfitting risk should have been successfully kept low in this work.

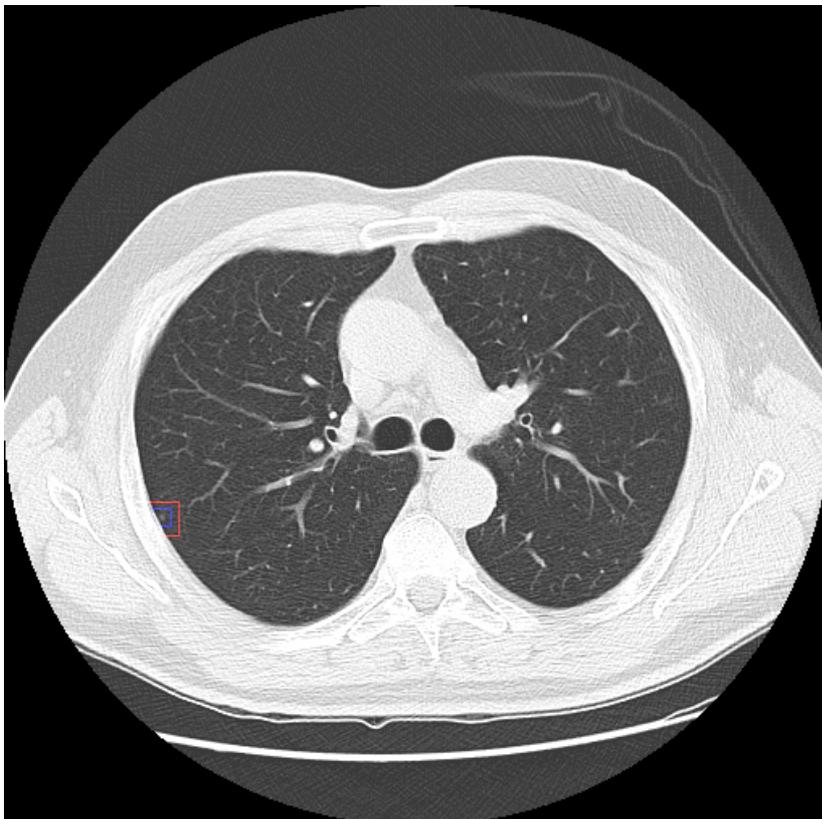


Figure 8: example of detected nodule (blue square: truth; red square: CAD detection) (DB5mm, patient 6-2-1-1, slice 37)

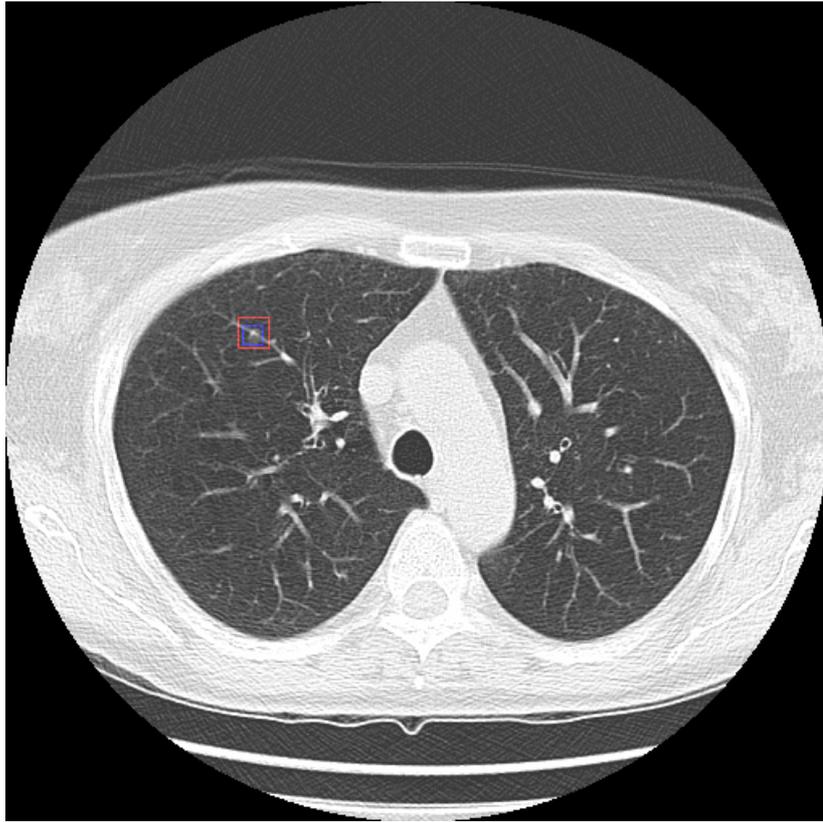


Figure 9: example of detected nodule (DB5mm, patient 9-1-1-1, slice 31)

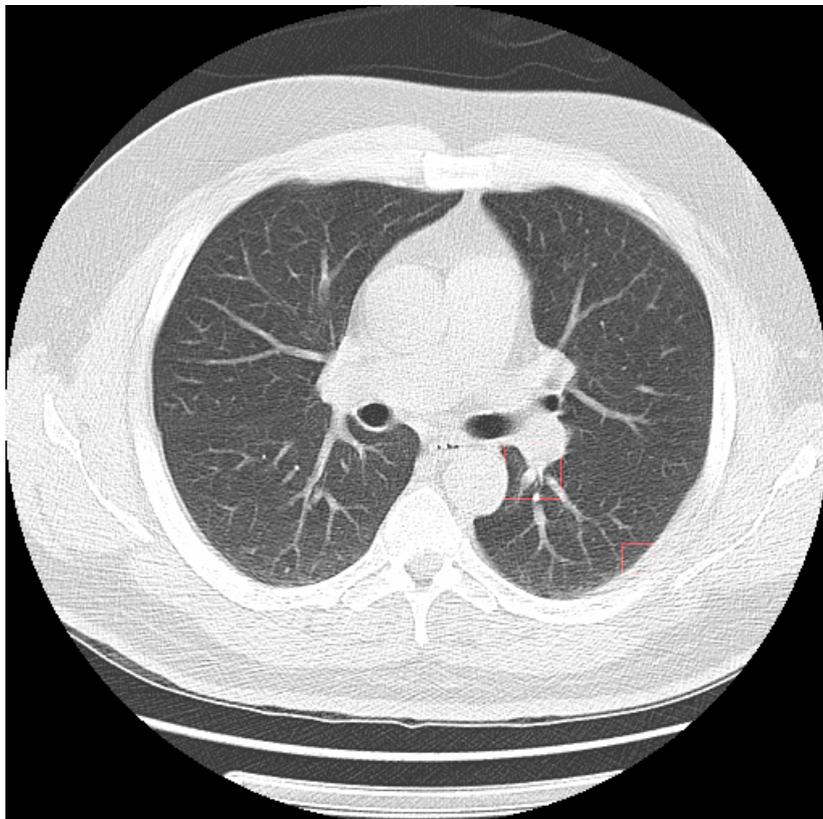


Figure 10: example of FPs of type *a* (right) and *b* (left) (DB5mm, patient 6-1-1-1, slice 35)

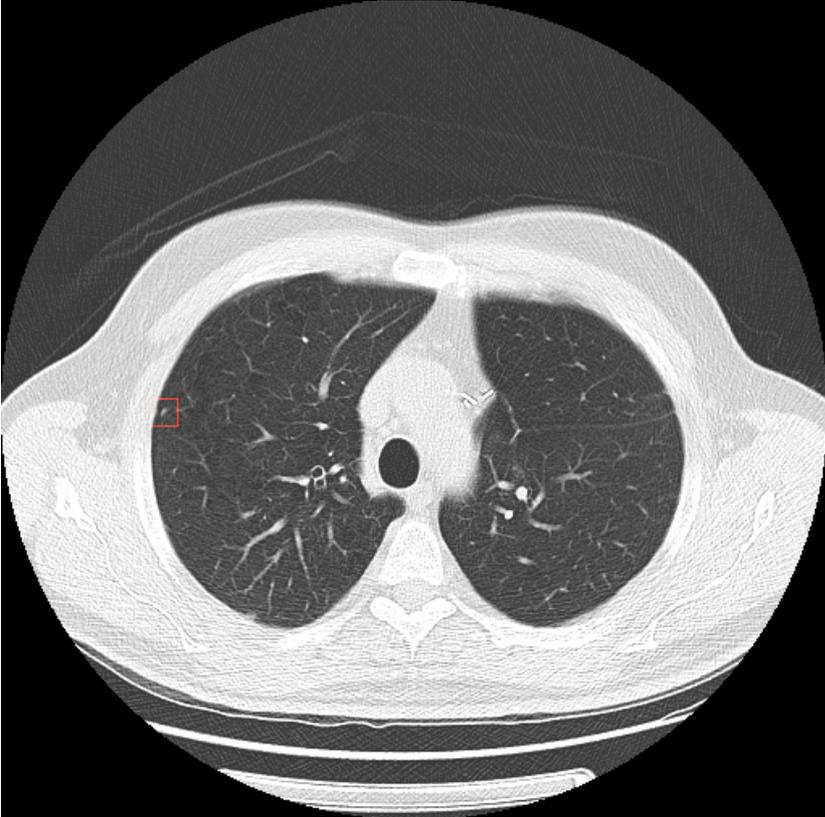


Figure 11: example of FP of type c (DB5mm, patient 7-1-1-1, slice 28)

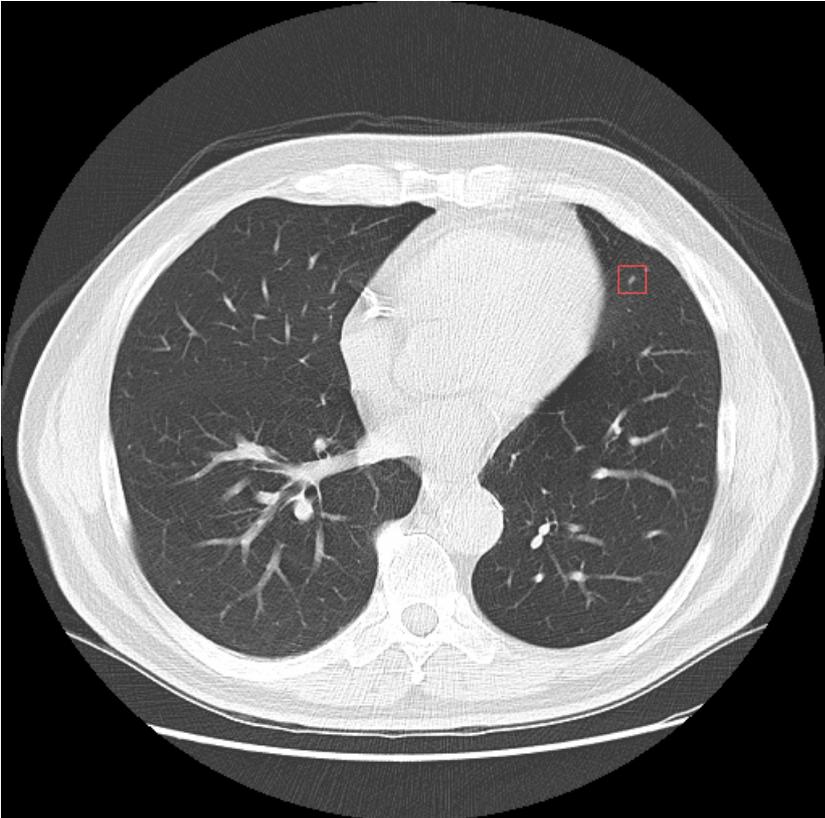


Figure 12: example of FP of type c (DB5mm, patient 7-1-1-1, slice 58)

Chapter 8

Analysis of LIDC database

The Lung Image Database Consortium (LIDC) is a cooperative enterprise of the National Cancer Institute (NCI) of the USA, aimed at developing an image database to serve as an international resource for the development and assessment of CAD systems applied to the detection of lung nodules in CT scans [Armato04].

Authors declare that the database will include:

- CT images and related technical data;
- nodule position and description;
- pathology information, when available.

Each CT scan will have been read by four radiologists two times, the first in blind mode, the second with knowledge of others results, and for each nodule larger than 3 mm a probabilistic description of contour will be provided, by taking into account results of the four readings, in order to capture inter-reader variability. Nodules smaller than 3 mm will only be described by their centroid.

8.1 Database at time of writing

At time of writing the database (DBLIDC) is composed of 23 nodules in 23 partial CT scans, with an average number of 38 slices per patient, slice thickness 0.625 mm and reconstruction increment 0.625 mm. Each image is 512 by 512 pixels and each pixel 0.703 mm in length. Minimum diameter of 2D nodule signals is 1.4 mm, maximum 58.9 mm. The nodule length in slice number is comprised between 5 and 43 slices.

By comparing DBLIDC with the database DB5mm, composed of slices of 5 mm thickness and 3 mm reconstruction increment, it is easy to understand that this database allows the detection of smaller nodules, and makes it easier as well: long objects are classified by means of more points of view (that is, ROIs).

Unfortunately, as already happened with both the other two databases used in this work, even this one is small, which means that the filtering and FPR-1 parameters have been chosen by means of reasonableness criteria, and that FPR-2 has undergone a Cross-Validation procedure, without independent validation. Noticeably, when the database will be enlarged, a complete validation procedure will be necessary for parameter optimization.

8.2 CAD processing I: detecting 2D signals

Threshold values (FRTHR) between 0.40 and 1.90 have been tested on the database, with tolerance for truth set, as usual, to 6.0 pixels. Truth was simply nodule centroid, taken from the probabilistic description given with each nodule. Searched nodules range was, as before, 2.4÷20.0 mm. Figure 1 below shows the number of FP/slice at varying FRTHR, both before and after the Scale Space Approach filtering.

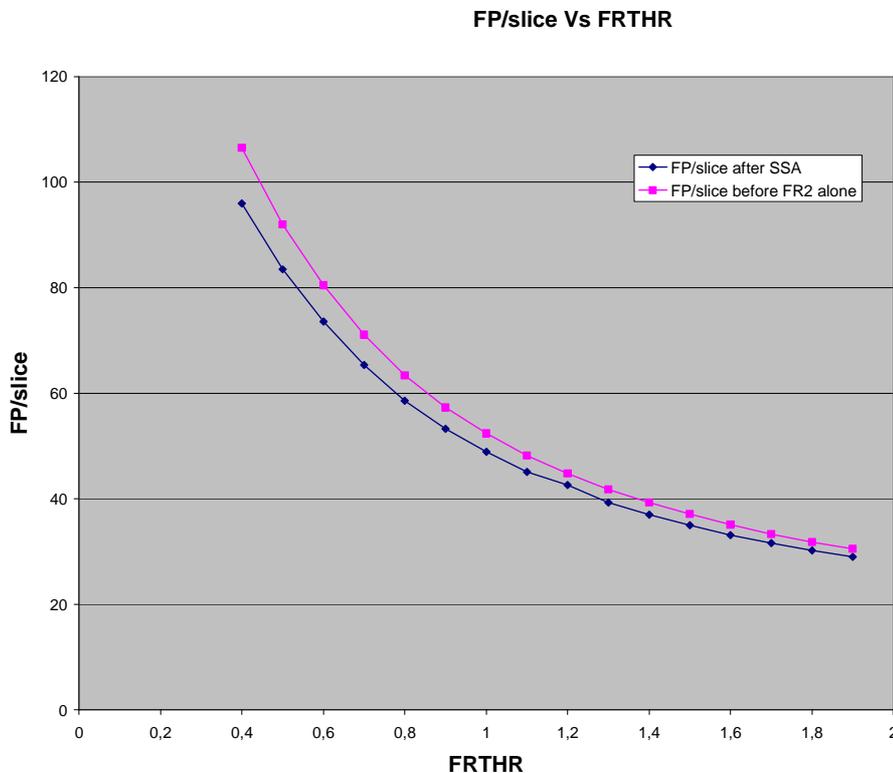


Figure 1: FP/slice at varying FRTHR before and after Scale Space filtering.

Because there was no validation set for this database, being the database itself too small, we decided to increase the value of $FRTHR$ until the average number of false positive signals per slice was similar to that of the DB5mm database for $FRTHR = 0.65$, the thresholding value considered in chapter 7. This choice is due the consideration that previous experience could be used as a guideline for new, unexplored databases, when no other landmarks are visible. With $FRTHR = 1.90$, the number of FP/slice was approximately 30, with 22 over 23 detected nodules, hence this value was chosen. Parameter *extraside* was set as for database DB5mm: 0.05.

8.3 CAD processing II: matching signals through slice and FPR-1

We grouped the 225 positive and 25,879 negative 2D signals across slices as described in chapter 6, with a tolerance between slices set to 7.75 pixels, and *Gap* to 1. Tolerance was slightly larger than before (7 pixels) because of the very irregular shapes of many nodules in this database, which could result in heavy nodule fragmentation across slices: such fragmentation was reduced to 6 unmatched positive signals and 2 split nodules. Since parameter *Gap* was not critical in database DB5mm, value 1 was chosen, being able to help reducing nodule fragmentation. In the end, the grouping process resulted in 30 3D positive signals (6 unmatched single signals, 24 non-singleton objects) and 4,625 3D negative signals (1.746 singletons).

After grouping, we determined FPR-1 cuts. As with database DB5mm, cuts were reasonably lax, in order to reduce to a minimum overfitting risk in such a small database. Figures 2 and 3 below show the distributions of the features: *theta*, *sigma-volume* and *length*.

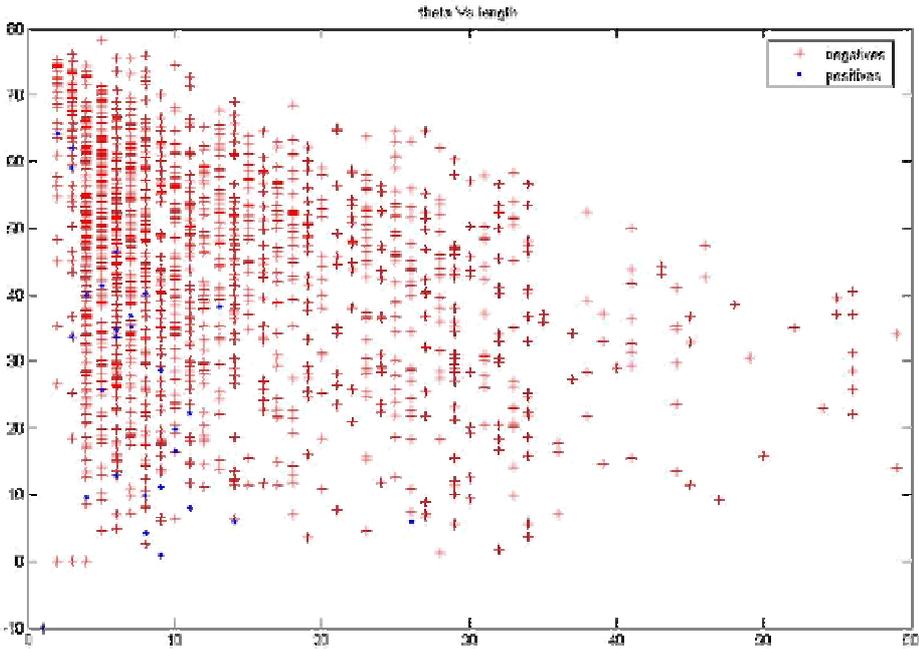


Figure 2: inclination - theta - in degrees Vs length of nodules (positives, blue), and not-nodules (negatives, red).

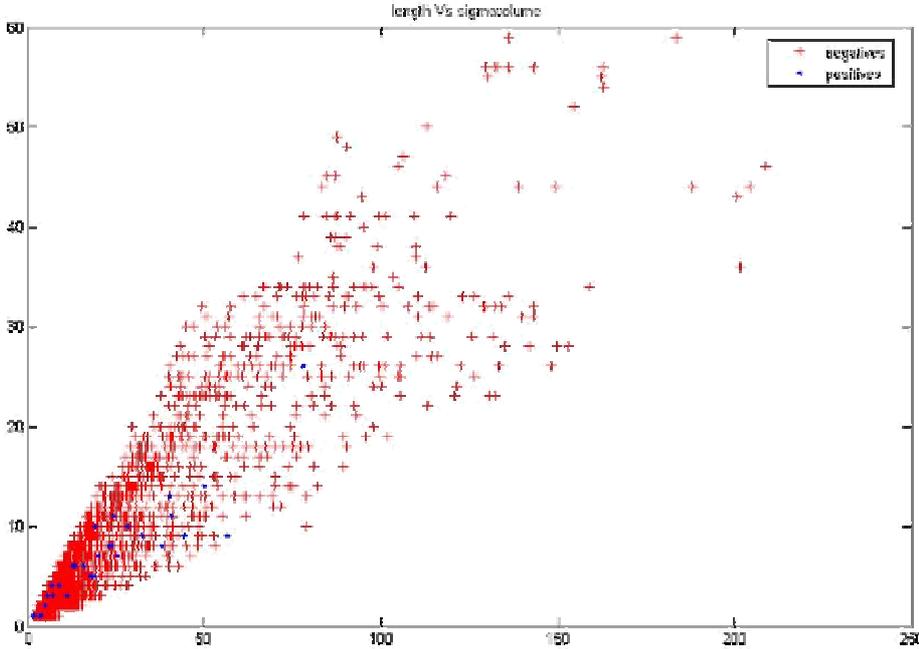


Figure 3: length Vs sigma-volume of nodules (positives, blue), and not-nodules (negatives, red).

Minimum length of non-singleton nodules was 3, maximum 26, and maximum length of negatives was 59. Maximum value of *sigma-volume* for nodules was 77, but

second maximum was 59, and all other values were under 50. Maximum value for theta was 64° for 2-slice long nodules, 38° for 13-slice long nodules, and 6° for 26-slice long nodules.

It was then decided to apply the following cuts:

1. *Length*: singletons and objects longer than 40 nodules;
2. *Theta*:
 - more than 70° for length 2 objects;
 - more than 65° for length 3 objects;
 - more than 60° for length 4 objects;
 - more than 55° for length 5 objects;
 - more than 50° for objects longer than 5;
3. *Sigma-volume*: greater than 80.

After the FPR-1 process, there remained 24 positive signals (2 of them were parts of other signals) and 2,132 negative signals (46% of the initial set). The only lost nodule was not detected by the FR^2 algorithm. No hybrid objects were found.

If we compare these figures with those of the DB5mm database, we see that after the present cuts a larger proportion of falses is still present: more than 40% instead of approximately 25%. It is hard to say, without a validation set, if it could be possible to reduce this proportion and still reach a good detection rate, so we decided not to change the cuts nor the detection threshold $FRTHR$, even if the initial aim to have a FP/slice figure close to the DB5mm one has not been satisfied after the FPR-1 phase. Our concerns were more focused on keeping overfitting risk as small as possible, than obtaining a small but less reliable FPs number.

8.4 CAD processing III: FPR-2

For what concerns the classification of 2D signals, we employed a Cross-Validation procedure for training and performance estimation, in particular a 7-Fold one, with 3 positives in the first 6 sub-sets, and 4 in the last one. As before with DB5mm, *RESIZE* was 11, 15, 19 or 23 pixels, *Gap* = 1 slices, SVM kernel: *polynomial*, with degree 2. Not as before, the training processes terminated with *almost* complete separability of the two classes, so we nonetheless decided not to change SVM *cost parameter* C , set equal to 100, and the SVM *unbalancing parameter* C^\pm , set equal to the ratio *number of negatives/ number of positives* in the training set. A few errors on a total

number of more than 10,000 vectors means that overlapping between the two classes is negligible hence could be ignored.

For what concerns 3D objects labelling, since nodules can be long more than 20 slices, we decided to use the MV algorithm with the same settings as in chapter 7, for objects long up to 8 slices (a 3D object is given a positive label if at least 1 signal is positive, in case it is long up to 4 signals, and if at least 2 signals are positives, in case it is between 5 and 8 slices), with the addition of these settings for longer objects:

- 3 (4) signals for $8 < \text{object length} \leq 12$;
- 5 (6) signals for $12 < \text{object length} \leq 16$;
- 8 (10) signals for $16 < \text{object length} \leq 20$;
- 10 (12) signals for $20 < \text{object length} \leq 30$;
- 12 (15) signals for objects longer than 30.

Numbers between brackets are other values that have been tested in different combinations with those chosen during the CV process, and later discarded. The minimum number of necessary positive 2D signals to give a positive label to a 3D object varies then between 20% and 40% of the number of signals in the object.

We tested the system with and without the Multirotation option. Results of the 7-Fold CV procedure with Multirotation only in training are compared with those without Multirotation in Table 1 below.

With Multirotation in training								
4-rotation threshold	2D pos.	2D neg.	3D pos. average	3D pos. overall average	3D neg.	FP/Pat. average	SVs number: positives	SVs number: negatives
At least 2	96-98 /191	2900- 2970 /13426	18/22	18/23	700- 720 /2132	30.8	~1800/4000	~2600/11500
	0.51	0.22	0.82	0.78	0.33		45%	23%
At least 3	65-74 /191	1520- 1680 /13426	15.2/22	15.2/23	360- 430 /2132	17.8	idem	idem
	0.45	0.12	0.69	0.66	0.19			
Without Multirotation in training								
At least 2	46-54 /191	1328- 1450 /13426	10.5/22	10.5/23	364- 393 /2132	16.6	~530/650	~1400/11500
	0.26	0.10	0.477	0.456	0.173		81%	12%

Table 1: comparisons between 7-Fold CV with and without Multirotation in training, both employing usual 4-rotation labelling in test, with “at least 2” and “at least 3” thresholds. The introduction of Multirotation has dramatically increased the 2D performance rate but also the 2D FP rate, in the case of “at least 2” labelling threshold. SVs are considered for an average training set of 6 subgroups of the total 7 ones.

The effect of Multirotation is a dramatic increase in both positives detection rate and FP/Patient. As already noticed with the DB5mm database, with threshold set as “at least 3” there is almost the same FP/Patient figure as without Multirotation, but with a much larger number of detected nodules. On the contrary, a noticeable difference with the DB5mm database is the absolute number of falses, both 2D and 3D: the percentage of the initial number here is approximately 10 times that of the previous database. The large number of falses and the heavy effect of Multirotation clearly show that the number of nodules in the database DBLIDC is at a critical low level. The number of SVs confirms this observation: almost half of positives and $\frac{1}{4}$ of negatives are SVs in Multirotation training, a situation very different from that with DB5mm database, where 33% of positives and only 5% of negatives were SVs.

Afterwards, Multirotation was considered for labelling. The following histograms (Figures 4 and 5) show the number of 2D signals that have received a positive label (data shown for *RESIZE* = 11).

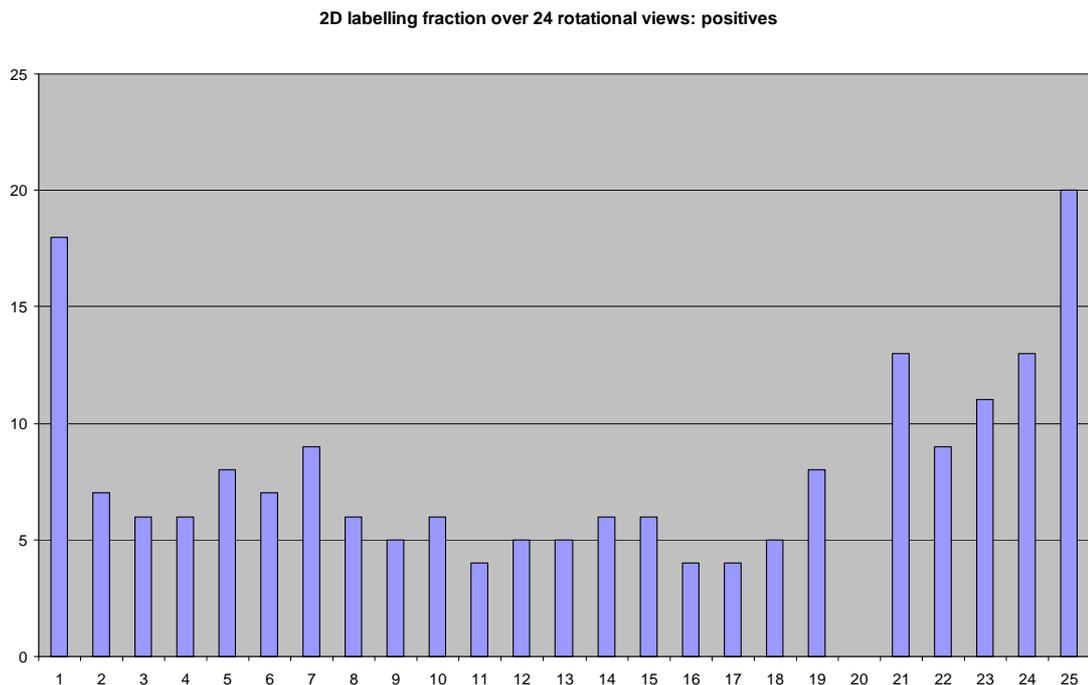


Figure 4: histogram shows how many 2D positive signals have received a certain number of positive labels, from 0 (column 1) to 24 (column 25).

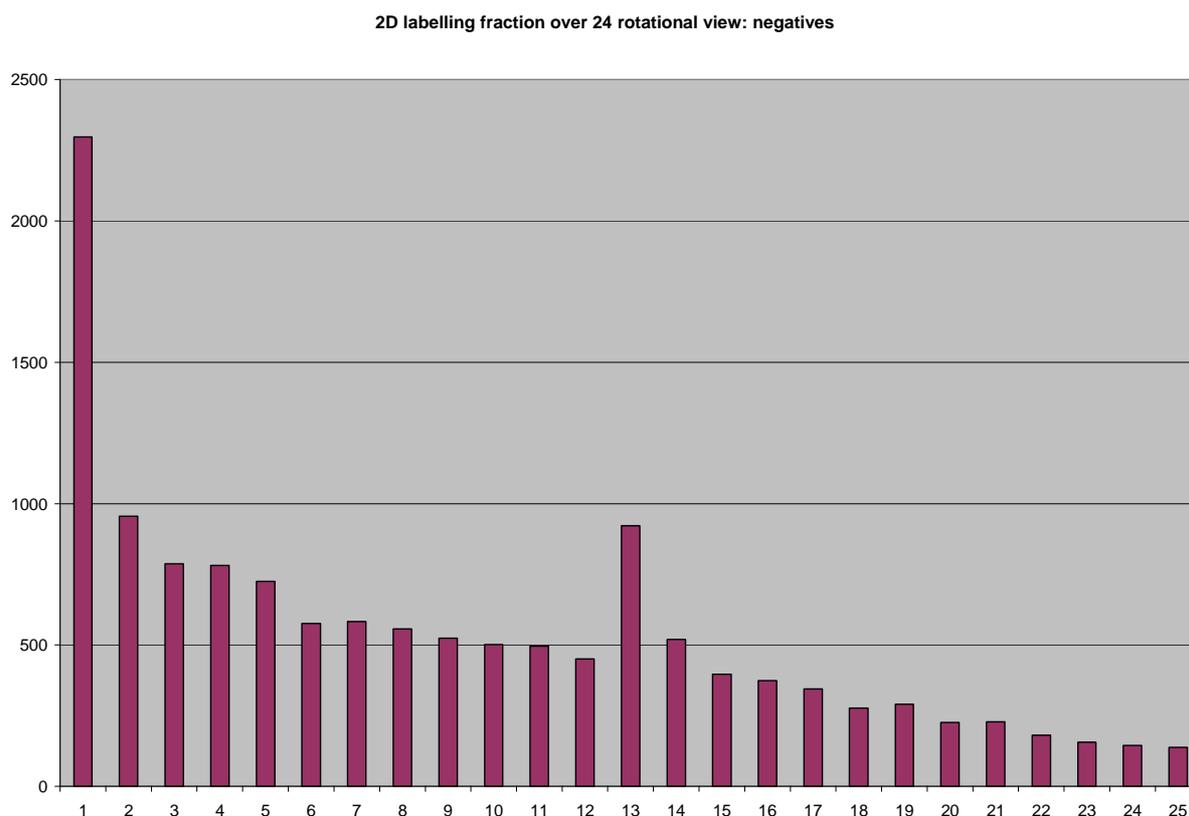


Figure 5: this histogram shows how many 2D negative signals have received a certain number of positive labels, from 0 (column 1) to 24 (column 25).

The ideal shape of the histograms would be that of a single column on the far right for the positive class and a single column on the far left for the other class. On the contrary, they both show shapes very different from the ideal ones, because the number of errors of the SVM classifier is very large, much larger than on the DB5mm database: there are 87 2D positives that receive no more than 11 positive labels over 24 (45% of 191), and 4,192 negatives that receive at least 12 positive labels (31% of 13,426). Besides, the two classes are not entirely separable in training, even if to an almost negligible degree (a few vectors over thousands), and are highly not separable in the test phase. This situation denotes that the two classes are not very well represented in the dataset, and in particular it is the nodule class that is underrepresented. This was among our expectations, because 23 nodules are a very small number, even to perform a Cross-Validation procedure. Figure 6 and 7 below show the effect of varying the labelling threshold from 12 to 20 over the positive and negative ROIs (all *RESIZE* values).

Correctly classified 2D positives at varying labelling threshold

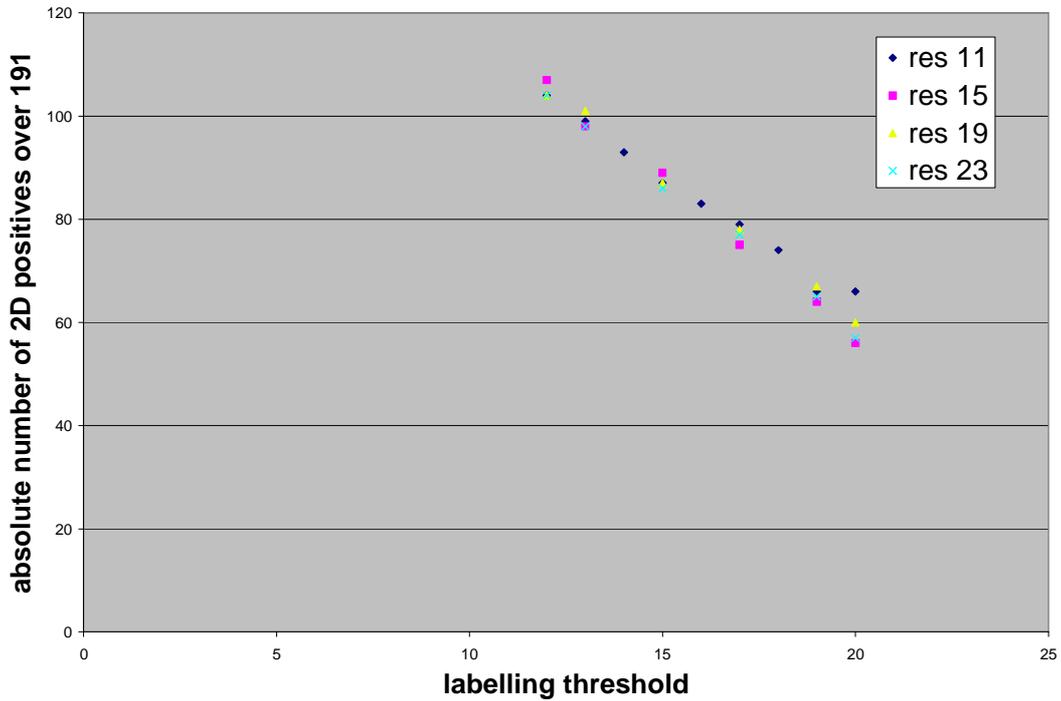


Figure 6: absolute number (over 191) of correctly classified 2D positive signals at varying labelling threshold from 12 to 24. In the legend, “res N” stays for “ROI resized to N pixel”.

Wrongly classified 2D negatives at varying labelling threshold

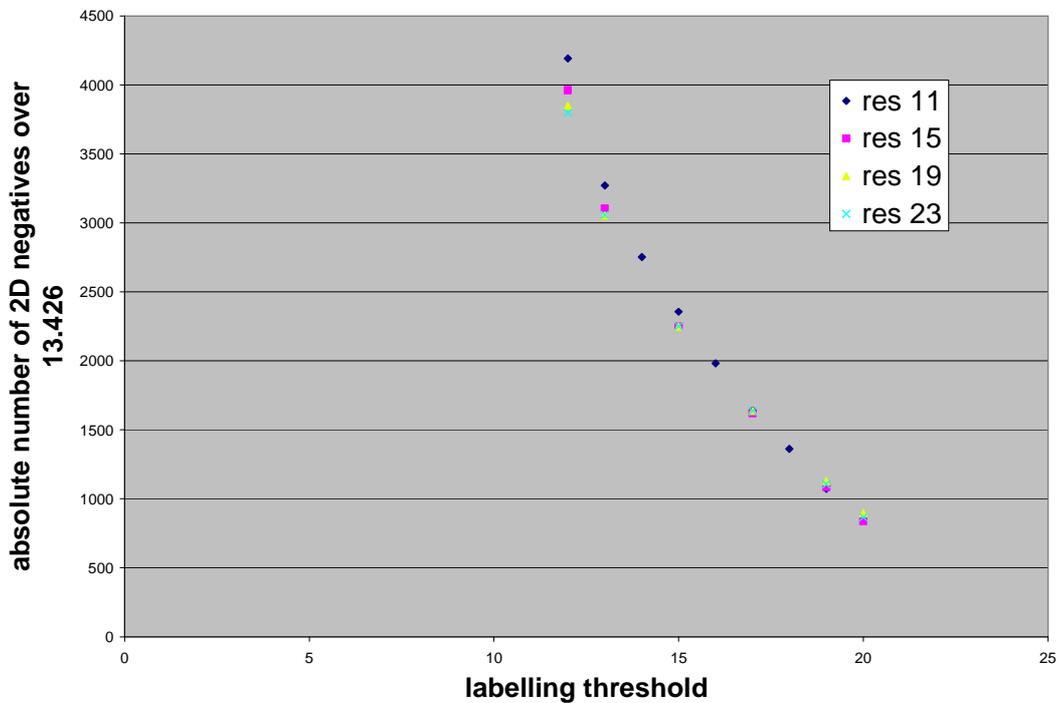


Figure 7: absolute number (over 13,426) of wrongly classified 2D negative signals at varying labelling threshold from 12 to 24. In the legend, “res N” stays for “ROI resized to N pixel”.

As before with DB5mm, the system behaviour seems in general to be not particularly dependent on the resizing parameter (*RESIZE*), except for the value $RESIZE = 11$, which presents a larger values of false positives for labelling threshold values from 12 to 15, and a larger values of true positives for threshold value equal to 20. Labelling thresholds over value 20 were not considered because of the already small performance rates.

As already noted, the smallness of the database makes difficult to clearly determine some fine details of the system behaviour, hence we decided to simply calculate its average behaviour over the *RESIZE* parameter to derive an FROC curve to summarize the system performances. In the table below (Table 2) we can find the performances of the system averaged over the resizing parameter.

2D labelling threshold (over 24)	12	13	15	17	19	20
2D pos	104,75	99	87,25	77,25	65,5	59,75
2D neg	3950	3121,75	2275,75	1633	1097,5	861,75
3D pos	19,5	18,5	16	13,75	12,25	10,75
3D pos: fraction	0,848	0,804	0,696	0,598	0,533	0,467
3D neg	956,75	738,5	533	383	260,5	210
2D FP/slice	4,907	3,878	2,827	2,029	1,363	1,07
3D FP/PAT	40,1	32,1	23,15	16,625	11,325	9,125

Table 2: absolute and fractional numbers of positive and negative signals detected by the system at varying labelling threshold. The database is formed by 191 2D positives, 13,426 2D negatives, 23 3D nodules and 23 patients (or cases). Results are averaged over the 4 different values of the *RESIZE* parameter. The number of 3D FP/Patient is calculated dividing the number of 3D negatives by the number of patients, whereas the number of 2D FP/slice by dividing the number of 2D falses by the number of patients times the average number of CT slices per patient (35).

The average percentage of correctly classified 2D positives varies from 31% to 55%, with an average percentage of wrongly classified 2D negatives varying between 6.4% and 29%. Correspondingly, 47% to 85% of nodules are detected, with 9.1 to 40.1 FP/Patient (or 1 to 5 FP/slice).

These averaged results allow us to plot the FROC curve of the system over 23 nodules in 23 patients of the database DBLIDC (see Figure 8 below).

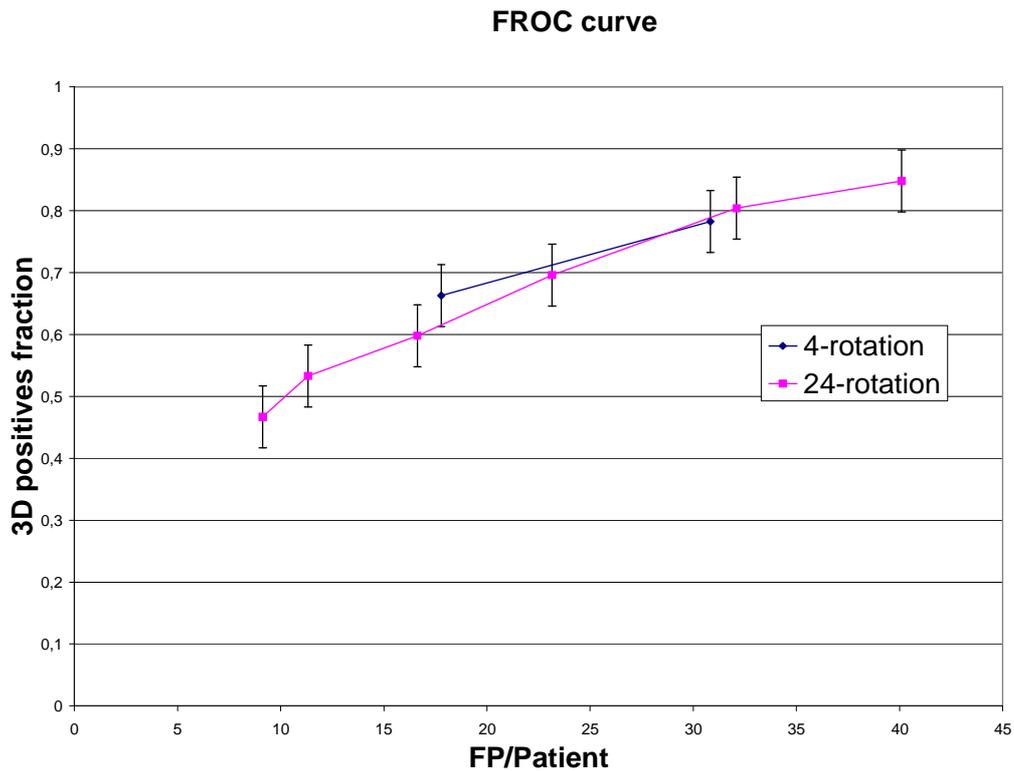


Figure 8: FROC curve of the CAD system for the DB5mm database. The 24-rotation curve refers to Multirotation algorithm used in test, with labelling threshold going from 12 to 20, whilst the 4-rotation curve refers to the usual 4-rotation algorithm, with labelling threshold equal to 2 and 3. In both, Multirotation procedure was employed in training. Error bars correspond to the standard deviation: 0.05.

Again, the FROC curve shows different working points of the system, which depend on the labelling threshold. As before, the 24-rotation curve allows a much finer tuning of the system than done by the 4-rotation curve, even if, this time, no clear difference in results can be spot, due to the small size of the database. With respect to the DB5mm database results, overall detection performances are comparable, but the number of FPs is generally larger: the FROC curve drops down faster when the FP/Patient figures diminishes. This reflects the fact that the SVM classifier makes a larger percentage of errors. In this case, the weakness of the small database plays a fundamental role.

In the same manner as before, errors bars in the FROC curve, about 0.05 in absolute value, approximately corresponding to 1 nodule, simply refers to the standard deviation of the 4 different curves corresponding to 4 different values of the *RESIZE* parameter. If we take into account the small database of only 23 nodules, we must conclude that larger errors, all originated from bad sampling of the database with respect to the *real*, and unknown, nodule distribution, might affect these results, and that these effects can't be directly estimated. As an alternative, we could consider again, as

in chapter 4 and 7, the binomial proportion interval estimation [Berger95][Brown02]: at 95% Confidence Interval, the error is approximately 4÷5 nodules (15÷20%).

8.5 False Positives analysis

The same three types of FPs found in database DB5mm are present here. Some examples of nodules and FPs are shown in the figures at the end of the chapter (*RESIZED* = 19, labelling threshold = 15).

8.6 Conclusions

LIDC database will for sure prove to be a great resource for the lung CAD community: unfortunately, at time of writing it is so small that only a partial validation of CAD systems is feasible. Because our CAD system needs to have many parameters optimized by means of a usual Estimation-Validation procedure, we initially considered the DB5mm database as a guideline and decided to set the FRTHR parameter in order to have a similar FP/slice value, then we choose FPR-1 step thresholds as much reasonable as possible, and finally we used a Cross-Validation procedure for training and performance estimation of FPR-2 step. With a large number of FP/Patient the overall CAD performances are similar to those of the CAD on the DB5mm database, but are worse when the FP/Patient rate moves towards zero. This variation is due to a performance decrease of both FPR-1 and FPR-2 steps, and the main reason is to be considered the narrowness of the database, which does not allow proper parameters estimation: with a large dataset parameters would be robust and their impact clearly estimated. Moreover, FPR-1 finds hard to eliminate falses also because of the use of partial CT scans: very long objects are truncated and made similar in length to nodules. Another important point is the quality and variety of 2D nodule views: even if in this database, with respect to the DB5mm one, there are more 2D nodule signals, many of them are very similar, because slices are very thin, so we can argue that the effective number of training sample is reduced; besides, the nodule class is wider, because there are many spiculated nodules. It must be said that we also tried to choose high-quality positive signals for the training phase as already done in chapter 7, by discarding approximately 30 signals over 191, but there were no appreciable differences in the final results, so we didn't report the details. Once more, the critical point is the database size.

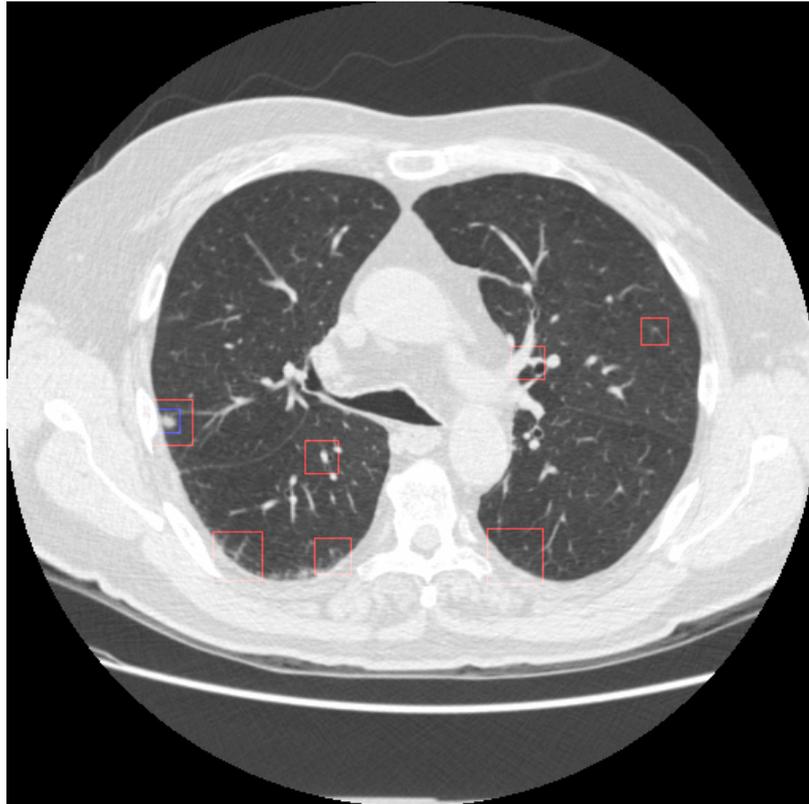


Figure 9: examples (blue square: truth; red square: CAD detection) of detected nodule and FPs. The three falses close to the bottom border are of type *a*, whilst the two inside lung area are of type *c*. (DBLIDC, patient 1, slice 23)

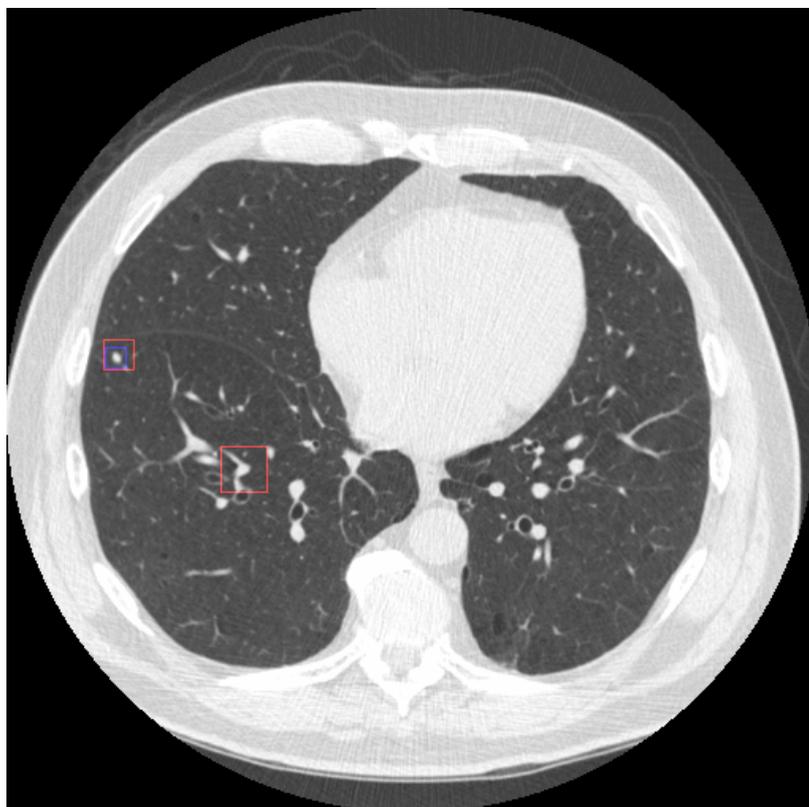


Figure 10: example of nodule and false (type *c*) (DBLIDC, patient 5, slice 21)

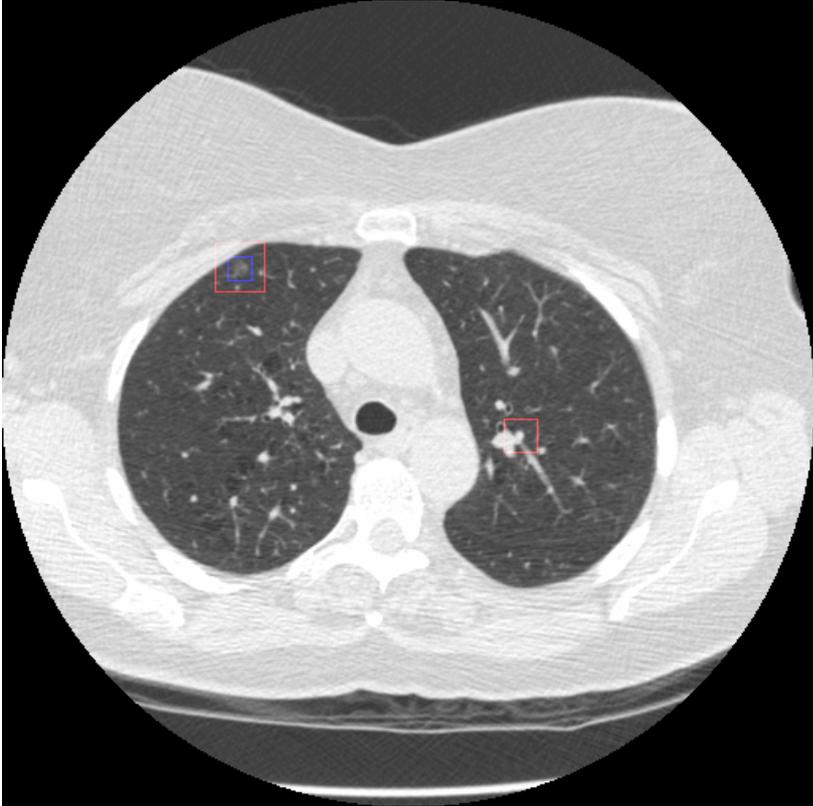


Figure 11: example of nodule and false (type *b/c*) (DBLIDC, patient 7, slice 15)

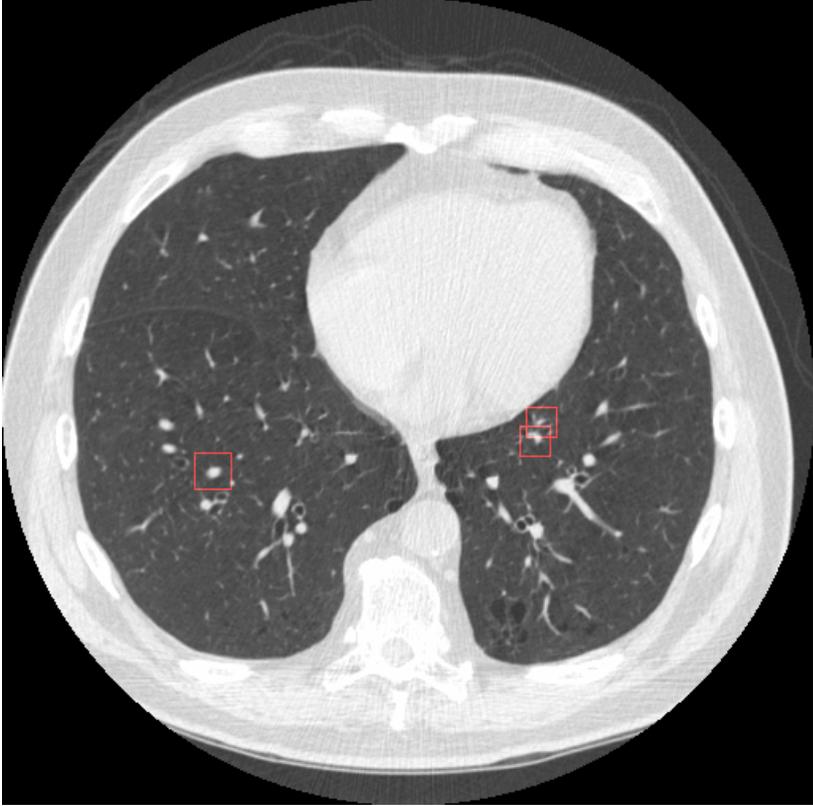


Figure 12: example of falses of type *c* (DBLIDC, patient 5, slice 6)

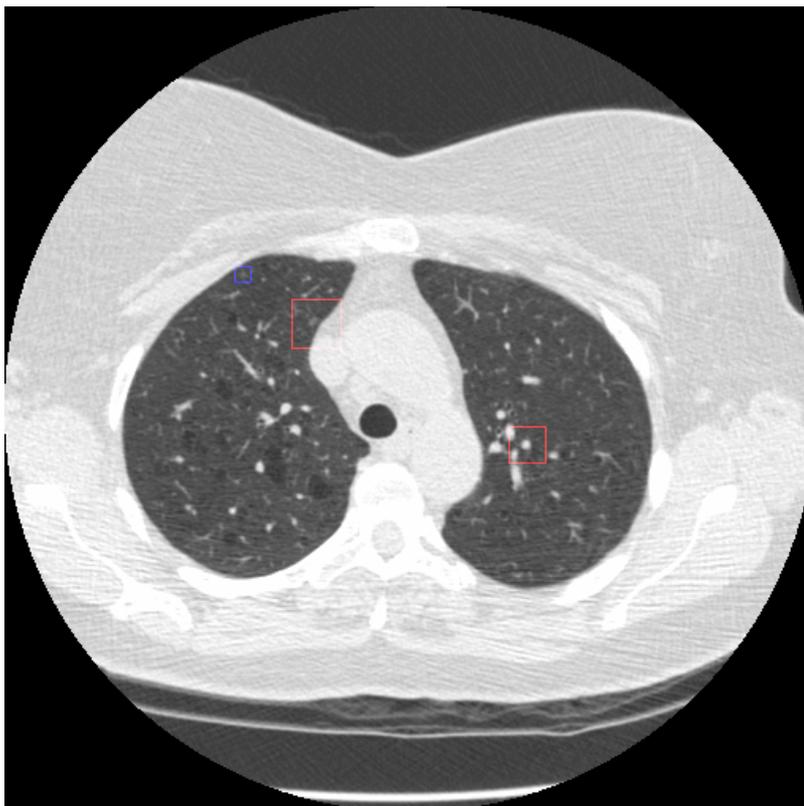


Figure 13: examples of a missed 2D nodule signal and of false positives of type a (left) and c (right) (DBLIDC, patient 7, slice 22)

Chapter 9

Conclusions

In this work we have described a new method for the detection of nodules in lung CT scans. At present, our CAD is composed of 5 macro steps:

1. Pre-processing :
 - a. DICOM to GL image transformation;
 - b. Automatic lung segmentation;
2. 2D candidate nodule signals detection: Fast Radial of Fast Radial filtering *AND* Scale Space filtering;
3. Grouping of 2D signals into 3D objects across slices;
4. False Positive Reduction , part 1 (coarse): simple cuts;
5. False Positive Reduction , part 2 (fine): SVM-based classification and labelling;

The segmentation part, which is completely automatic, reduces the area to be subsequently analyzed by the detection system, thus sinking the processing times of subsequent steps and eliminating false positive signals.

The filtering approach is completely new: in particular, the iterative filtering with the Fast Radial Filter is original, as well as the innovative and effective use of the resonance given by the Normalized Laplacian of the Scale Space Approach. This resonance has some very desirable properties: the experimentally proved relationship between circular objects' size and the resonance sigma value, the fine and easy control of nodule size and position.

The abandoning of the Seeded Region Growing algorithm and its unstable results is another important consequence of the SSA. The only weak point of SSA is a border resonance effect (*partial border resonance*), which gives rise to a subgroup of false positive signals, which will require *ad hoc* treatment in future.

The simple and efficient FPR-1 phase is also partly based on the SSA.

The FPR-2 phase employing Gray Level features and an SVM classifier is quite promising, especially used in conjunction with the Multi-rotation approach both during training and during test phases. During the latter, the Multi-rotation approach results in a finer tuning of the CAD working points (FROC curve).

A big difficulty we had to face during the CAD development was the lack of a large database: a very common problem among the CAD community that will hopefully be resolved by the LIDC enterprise, unfortunately still at its beginning.

Taking into account lung CAD literature, we can see that the work described in this thesis has reached satisfactory results, comparable to those of other larger research groups; nonetheless, it is clear that much effort is still necessary to bring this CAD to its full development. Some of the possible future improvements are, to name but a few: the employment of different ROI features, an FPR-3 step based on SVR-filtering, an *ad hoc* post-processing to get rid of some false positive signals, a different FPR-2 step based on an ensemble of classifiers trained with different positives or false subgroups. Taking into account the rapid technological advances of CT, constantly moving towards thinner and thinner slices, 3D interpolation of signals will probably become very interesting for nodule detection.

Glossary

ANN: Artificial Neural Network
CAD: Computer Aided Detection
CT: Computed Tomography
CV: Cross Validation
CXR: Chest X-Ray
DB: Data Base
DICOM: Digital Imaging and Communication in Medicine
DOG: Difference of Gaussians
FP: False Positive
FPF: False-Positive Fraction
FP/slice: False Positive per slice
FPR: False-Positive Reduction
FR: Fast Radial
FR²: Fast Radial of Fast Radial
FROC: Free Response Operating Characteristic
FRTHR: FR or FR² Threshold
GATM: Genetic Algorithm Template Matching
GE: General Electric
GL: Gray Level
HU: Hounsfield Unit
IEO: Istituto Europeo di Oncologia
LDA: Linear Discriminant Analysis
LIDC: Lung Image Database Consortium
LOO: Leave One Out
MF: Matching Filter
MTANN: Massive Training Artificial Neural Network
MV: Majority Voting
NEMA: National Electrical Manufacturers Association
NL: Normalized Laplacian
NLO: Noise Lowered FR

Glossary

ORBA: Orientation Based FR

PACS: Picture Archiving and Communication Systems

RBF: Radial Basis Function

ROC: Receiver Operating Characteristic

ROI: Region Of Interest

SRG: Seeded Region Growing

SSA: Scale Space Analysis

SV: Support Vector

SVC: Support Vector Clustering

SVM: Support Vector Machine

SVR: Support Vector Regression

TPF: True Positive Fraction

WL: Window Level

WW: Window Width

References

- [AACR05]: American Association for Cancer Research, <http://www.aacr.org>
- [AIRC06]: <http://www.airc.it>
- [Adams94]: R. Adams and L. Bishof, "Seeded Region Growing", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16, 6, 641-647, 1994
- [Angelini05]: E. Angelini, R. Campanini and A. Riccardi, "Support vector regression filtering for reduction of false positives in a mass detection cad scheme: preliminary result", University of Bologna Preprint, 2005. Available at: <http://amsacta.cib.unibo.it/archive/00000912/>
- [Armato99a]: S. G. Armato III, M. L. Geiger, C. J. Moran, J. T. Blackburn, K. Doi, and H. MacMahon, "Computerized detection of pulmonary nodules on CT scans", *Radiographics* 19, 1303-1311, 1999
- [Armato99b]: S. G. Armato III, M. L. Geiger, J. T. Blackburn, K. Doi, and H. MacMahon, "Three-dimensional approach to lung nodule detection in helical CT", *Proc SPIE* 3661, 553-559, 1999
- [Armato00]: S. G. Armato III, M. L. Geiger, and H. MacMahon, "Analysis of a three-dimensional lung nodule detection method for thoracic CT scans", *Proc SPIE* 3979, 103-109, 2000
- [Armato01]: S. G. Armato III, M. L. Geiger, K. Doi, U. Bick, and H. MacMahon, "Computerized lung nodule detection: comparison of performance for low-dose and standard-dose helical CT scans", *Proc SPIE* 4322, 1449-1454, 2001
- [Armato02]: S. G. Armato III, F. Li, M. L. Geiger, H. MacMahon, S. Sone, K. Doi, "Lung cancer: performance of automated lung nodule detection applied to cancers missed in a CT screening program", *Radiology* 225, 685-692, 2002
- [Armato03]: S. G. Armato III, M. B. Altman, and P. J. La Riviere, "Automated detection of lung nodules in CT scans: effect of image reconstruction", *Med. Phys.* 30 (3), 461-472, March 2003

References

- [Armato04]: S. G. Armato III, G. McLennan, M. F. McNitt-Gray, C. R. Meyer, D. Yankelevitz, D. R. Aberle, C. I. Henschke, E. A. Hoffman, E. A. Kazerooni, H. McMahon, A. P. Reeves, B. Y. Croft, L. P. Clarke, “Lung Image Database Consortium: developing a resource for the medical imaging research community”, *Radiology* 2004, 232: 739-748
- [Austin96]: J.H. Austin, N.L. Mueller, P.J. Friedman, et al., “Glossary of terms for CT of the lungs: recommendation of the Nomenclature Committee of the Fleischner Society”, *Radiology* 1996, 200:327-331
- [Bazzani01]: A. Bazzani, A. Bevilacqua, D. Bollini, R. Brancaccio, R. Campanini, N. Lanconelli, A. Riccardi, D. Romani, “An SVM classifier to separate false signals from microcalcifications in digital mammograms”, *Physics in Medicine and Biology*, Vol 46(6), June 2001, 1651-1663
- [Bechtel03]: J. J. Bechtel and T. L. Petty, “Strategies in lung cancer detection: achieving early identification in patients at high risk”, *Postgrad Med* 2003; 114(2): 20-6
- [Bellomi04]: Prof M. Bellomi’s personal communication, 2004
- [Berger95]: J. O. Berger, “Statistical Decision Theory and Bayesian analysis, 2nd Ed.”, Springer-Verlag, New York 1995
- [Boyle05]: P. Boyle and J. Ferlay, “Cancer incidence and mortality in Europe, 2004”, *Annals of Oncology* 2005 16(3):481-488
- [Brenner04]: D. Brenner, “Radiation risks potentially associated with low-dose CT screening of adult smokers for lung cancer”, *Radiology* 2004, 231: 440-445
- [Brown97]: M. S. Brown, M. F. McNitt-Gray, N. J. Mankovich, J. G. Goldin, J. Hiller, L. S. Wilson, and D. Aberle, “Method for Segmenting Chest CT Image Data Using an Anatomical Model: Preliminary Results”, *IEEE Transactions on Medical Imaging*, 16, 6:828-839, 1997
- [Brown02]: L. D. Brown, T. T. Cai and A. DasGupta, “Confidence intervals for a binomial proportion and asymptotic expansions”, *The Annals of Statistics* 30, 1, 160-201, 2002

- [Brown03]: M. S. Brown, J. G. Goldin, R. D. Suh, M. F. McNitt-Gray, J. W. Sayre, D. R. Aberle, “Lung micronodules: automated method for detection at thin-section CT – initial experience”, *Radiology* 226, 256-262, 2003
- [Campanini04]: R. Campanini, D. Dongiovanni, E. Iampieri, N. Lanconelli, M. Masotti, G. Palermo, A. Riccardi and M. Roffilli, “A novel featureless approach to mass detection in digital mammograms based on Support Vector Machines”, *Physics in Medicine and Biology*, IOP Publishing, UK, 49, pp. 961-975, 2004
- [Campanini06]: R. Campanini’s personal communication
- [Castleman96]: K. Castleman, “Digital Image Processing”, Prentice Hall, 1996
- [Chang04]: F. Chang, C.-J. Chen, and C.-J. Lu, “A Linear-Time Component-Labeling Algorithm Using Contour Tracing Technique”, *Computer Vision Image Understanding*, 93(2):206--220, 2004
- [DICOM06]: <http://medical.nema.org/>, <http://dicom.offis.de/dcmthk.php.en>
- [Di Gesù95]: V. Di Gesù and C. Valenti, “The Discrete Symmetry Transform in Computer Vision”, Technical Report DMA 011 95, Palermo University, 1995
- [Dodd04]: Lori E. Dodd, PhD, Robert F. Wagner, PhD, Samuel G. Armato III, PhD, Michael F. McNitt-Gray, PhD, Sergey Beiden, PhD Heang-Ping Chan, PhD, David Gur, ScD, Geoffrey McLennan, MD, PhD, Charles E. Metz, PhD, Nicholas Petrick, PhD Berkman Sahiner, PhD, Jim Sayre, Dr.PH, the Lung Image Database Consortium Research Group, “Assessment Methodologies and Statistical Issues for Computer-Aided Diagnosis of Lung Nodules in Computed Tomography: Contemporary Research Topics Relevant to the Lung Image Database Consortium”, *Acad Radiol* 2004; 11:462– 475
- [Duda00]: R. O. Duda, P. E. Hart and D. G. Stork, “Pattern Classification (2nd Edition)”, Wiley Interscience, 2000.
- [Efron83]: B. Efron, “Estimating the error rate of a prediction rule: improvement on Cross-Validation”, *J. Am. Stat. Ass.* 78(382), 316-330, 1983
- [El-Naqa02]: I. El-Naqa, Y. Yang, M. N. Wernick, N. P. Galatsanos and R. M. Nishikawa, “A Support Vector Machine approach for detection of microcalcifications”, *IEEE Transactions on Medical Imaging* 21:1552-63, 2002

References

- [Erasmus00]: J. F. Erasmus, H. P. MacAdams, J. E. Connolly, “Solitary pulmonary nodules: part II. Evaluation of the indeterminate nodule”, *Radiographs* 2000; 20:59-66
- [Fukano03]: G. Fukano, H. Takizawa, K. Shigemoto, S. Yamamoto, T. Matsumoto, Y. Tateno, T. Inuma, “Recognition method of lung nodules using blood vessel extraction techniques and 3D object models”, *PROC. SPIE* 5032, 190-198, 2003
- [Ge05]: Z. Ge, B. Sahiner, H. Chan, L. M. Hadjiiski, P. N. Cascade, N. Bogot, E. Kazerooni, J. Wei, and C. Zhou, “Computer-aided detection of lung nodules: False positive reduction using a 3D gradient field method and 3D ellipsoid fitting”, *Med. Phys.* 32 (8), 2005, 2443-2454.
- [Geman92]: S. Geman, E. Bienenstock, R. Doursat, “Neural Networks and the Bias/Variance Dilemma”, *Neural Computation* 4, 1-58, 1992
- [Gonzalez02]: R. C. Gonzalez, R. E. Woods, “Digital Image Processing, Second Edition”, Prentice Hall 2002
- [Goo05]: J. Mo Goo, “Computer Aided Detection of Lung Nodules on Chest CT: Issues to be solved before Clinical Use”, *Korean J. Radiol* 2005, 6(2):62–63
- [Gur04]: D. Gur, B. Zheng, C. R. Fuhrman, L. Hardesty, “On the testing and reporting of Computer-Aided detection Results for Lung Cancer Detection”, *Radiology* 2004, 232:5–6
- [Gurcan02]: 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”, *Med. Phys.* 29 (11), 2552-2558, November 2002
- [Haykin99]: S. Haykin, “Neural Networks, a comprehensive foundation”, Prentice Hall 1999
- [Hastie01]: T. Hastie, R. Tibshirani, J. H. Friedman: “The elements of Statistical Learning”, Springer 2001
- [Henschke99]: C. L. Henschke, D. I. MacCauley, D. F. Yankelevitz, et al., “Early Lung Cancer Action Project: overall design and findings from baseline screening”, *Lancet* 1999; 354:99-105

- [Henschke04]: C. I. Henschke, D. F. Yankelevitz, D. P. Naidich, D. I. MacCauley, G. MacGuinness, D. M. Libby, J. P. Smith, M. W. Pasmantier, O. S. Miettinen, "CT screening for lung cancer: suspiciousness of nodules according to size on baseline scans", *Radiology* 231, 164-168, 2004
- [Hu01]: S. Hu, E. A. Hoffmann, and J. M. Reinhardt, "Automatic Lung Segmentation for Accurate Quantitation of Volumetric X-Ray CT Images", *IEEE Transactions on Medical Imaging*, 20, 8:490-498, 2001
- [Huang79]: T.S. Huang, G.J. Yang, and G.Y. Tang, "A Fast Two-Dimensional Median Filtering Algorithm", *IEEE Trans. Acoustics Speech Signal Proc.*, vol. ASSP-27, pp. 13-18, Feb. 1979
- [Kalendar00]: W. A. Kalendar, "Computed Tomography", Wiley-VCH 2000
- [Kapur85]: Kapur J.N., Sahoo P.K., and Wong A.K.C., "A new method for grey-level picture thresholding using the entropy of the histogram", *Graphical Models and Image Processing*, 29: 273-285, 1985
- [Kaucic03]: R. A. Kaucic, C. C. McCulloch, P. R. S. Mendonça, D. J. Walter, R. S. Avila, J. L. Mundy, "Model-based detection of lung lesions in CT exams", *Proceedings of the International Conference on Computer Assisted Radiography and Surgery CARS'03*, Elsevier International Congress Series 1256, 990-997, 2003
- [Kittler98]: J. Kittler, M. Hatef, R. P. W. Duin and J. Matas, "On Combining Classifiers", *IEEE TPAAMI*, 20, 3:226-239, 1998
- [Koenderink84]: J. J. Koenderink, "The structure of images", *Biol. Cyb.*, 50. 363-370, 1984
- [Kohavi95]: R. Kohavi, "A study of Cross-Validation and Bootstrap for accuracy estimation and model selection", *International Joint Conference on Artificial Intelligence (IJCAI) 1995*
- [Koike99]: T. Koike, M. Terashima, T. Takizawa, et al., "The influence of lung cancer mass in surgical results", *Lung Cancer* 1999; 24(2): 75-80
- [Kovesi97]: P. D. Kovesi, "Symmetry and Asymmetry from Local Phase", *Proc. 10th Australian Joint Conference on Artificial Intelligence*, 1997

References

- [Kovesi05a]: P. D. Kovesi, MATLAB and Octave Functions for Computer Vision and Image Processing, School of Computer Science & Software Engineering, The University of Western Australia. Available from: <http://www.csse.uwa.edu.au/~pk/research/matlabfns/>
- [Kovesi05b]: P. D. Kovesi's personal communication, 2005
- [Lee01]: 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
- [Lee05]: J. Lee, G. Gamsu, J. Czum, N. Wu, R. Johnson, S. Chakrapani, "Lung Nodule detection on chest CT: evaluation of a Computer Aided detection (CAD) system", *Koran J. Radiol* 2005, 6: 89–93
- [LIDC03]: LIDC education exhibit, *RSNA 2003*
- [Li03]: Q. Li, S. Sone, K. Doi, "Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans", *Med. Phys.* 30 (8), 2040-2051, August 2003
- [Li05a]: Q. Li, F. Li, K. Doi, "Computerized nodule detection in thin-slice CT using selective enhancement filter and automated rule-based classifier", *Proc. SPIE 2005, Medical Imaging*, 5747, 104-112
- [Li05b]: F. Li, H. Arimura, K. Suzuki, J. Shiraishi, Q. Li, H. Abe, R. Engelmann, S. Done, H. MacMahon, and K. Doi, "Computer-aided Detection of Peripheral Lung Cancers missed at CT: ROC analyses without and with localization", *Radiology* 2005, 237: 684-690
- [Lindeberg93]: T. Lindeberg, "On Scale Selection for Differential Operators", *Proc. 8th Scandinavian Conference on Image Analysis*, May 1993, 857-866
- [Lowe04]: D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", *International Journal of Computer Vision*, 60(2), 91–110, 2004
- [Loy03]: G. L. Loy and A. Zelinsky, "Fast Radial Symmetry for Detecting Points of Interest", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25, 8, 2003

- [Masotti05]: M. Masotti, "Optimal image representations for mass detection in digital mammography", PhD thesis, Physics Department, University of Bologna, 2005. Available at <http://amsacta.cib.unibo.it/archive/00001090/>
- [Mehnert97]: A. Mehnert and P. Jackway, "An improved seeded region growing algorithm", *Pattern Recognition Letters* 18 (1997), 1065-1071
- [Metz86]: C. E. Metz, "ROC Methodology in Radiologic Imaging", *Investigative Radiology* 1986; 21:720-733
- [NLHEP06]: <http://www.nlhep.org/>
- [Nagao79]: M. Nagao and T. Matsuyama, "Edge preserving smoothing", *Computer Graphics and Image Processing*, vol. 9, pp. 394-407, Apr. 1979
- [Otsu79]: Otsu N., "A threshold selection method from gray level histograms", *IEEE Trans. on Systems, Man and Cybernetics*, 9: 62-66, 1979
- [Paik05]: D. S. Paik, C. F. Beaulieu, G. D. Rubin, B. Acar, R. Brooke 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 Transactions on Medical Imaging* 2004, 23, 6:661-675
- [Patz04]: Edward F. Patz, Jr, Stephen J. Swensen, James E. Herndon, II, "Estimate of Lung Cancer Mortality from Low-Dose Spiral Computed Tomography Screening Trials: Implications for Current Mass Screening Recommendations", *Journal of Clinical Oncology*, Vol 22, No 11 (June 1), 2004: pp. 2202-2206
- [Preteux89]: F. Pretoux, N. Merlet, P. Grenier, and M. Mouehli, "Algorithms for automated evaluation of pulmonary lesions by high resolution CT via image analysis", in *Proc. Radiol. Soc. Of North America RSNA'89*, p.416, 1989
- [Rao94]: K. R. Rao and J. Ben-Arie, "Multiple Template Matching using Expansion Matching," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 4, No. 5, pp. 490-503, October 1994
- [Reisfield95]: D. Reisfield, H. Wolfson and Y. Yeshurun, "Context free attentional operators: the generalized symmetry transform", *International Journal of Computer Vision*, special issue on qualitative vision, 14, 119-130, 1995

References

- [Riccardi00]: A. Riccardi, “Classificazione di immagini mammografiche con il metodo SVM”, Tesi di Laurea in Fisica-Laurea degree dissertation, Università di Bologna, 2000
- [Ridler78]: T.W. Ridler, S. Calvard, “Picture thresholding using an iterative selection method”, *IEEE Trans. System, Man and Cybernetics*, SMC-8 (1978) 630-632.
- [Rubin05]: G. D. Rubin, J. K. Lyo, D. S. Paik, A. J. Sherbondy, L. C. Chow, A. N. Leung, R. Mindelzun, P. K. Schraedley-Desmond, S. E. Zinck, D. P. Naidich, S. Naipel, “Pulmonary nodules on Multi-Detector Row CT Scans: Performance Comparison of Radiologists and Computer-aided Detection”, *Radiology* 2005; 234:274-283
- [Soille96]: Jones, Ronald, and P. Soille, “Periodic lines: Definition, cascades, and application to granulometrie”, *Pattern Recognition Letters*, Vol. 17, 1996, pp. 1057-1063
- [Soille99]: P. Soille, “Morphological Image Analysis – Principles and Applications”, *Springer-Verlag*, 1999
- [Suzuki03]: K. Suzuki, S. G. Armato III, F. Li, S. Sone, and K. Doi, “Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography”, *Med. Phys.* 30 (7), 1602-1617, July 2003
- [Swensen05]: Stephen J. Swensen, James R. Jett, Thomas E. Hartman, David E. Midthun, Sumithra J. Mandrekar, Shauna L. Hillman, Anne-Marie Sykes, Gregory L. Aughenbaugh, Aaron O. Bungum and Katie L. Allen, “CT Screening for Lung Cancer: Five-year Prospective Experience”, *Radiology* 2005, 235: 259 - 265
- [Tomita77]: F. Tomita and S. Tsuji, “Extraction of multiple regions by smoothing in selected neighborhoods”, *IEEE Trans. Systems, Man, and Cybernetics*, vol. SMC-7, pp. 107-109, Feb. 1977
- [Yamamoto96]: S. Yamamoto, M. Matsumoto, Y. Tateno, T. Iinuma, and T. Matsumoto, “Quoit filter: a new filter based on mathematical morphology to extract the isolated shadow, and its application to automatic detection of lung cancer in X-ray CT”, *Proc. 13th Int. Conf. Pattern Recognition* 2, 3-7, 1996
- [Vapnik95]: V.N. Vapnik, “The Nature of Statistical Learning Theory”, Springer 1995
- [Vapnik98]: V. N. Vapnik, “Statistical Learning Theory”, J. Wiley 1998

[Wiemker02]: R. Wiemker, P. Rogalla, A. Zwartkruis, T. Blaffer, “Computer Aided lung nodule detection on high resolution CT data”, *Proc. SPIE 2002, Medical Imaging*, 4684, 677-688

[Witkin83]: A. P. Witkin, “Scale Space filtering”, *Proc. 8th IJCAI*, Karlsruhe, Germany, 1983, 1019-1022

[Wu05]: K. Wu, E. Otoo, A. Shoshani, “Optimizing Connected Component Labelling Algorithms”, *Proceedings of SPIE Medical Imaging Conference 2005*

References

Acknowledgements

Many thanks to the members of the Medical Imaging Group in Bologna, in particular to its chief prof. Campanini, and especially to Todor, with whom I have worked in the last two years, and who freely and successfully developed in C language the pre-processing algorithm, starting from the initial Matlab version.

Many thanks to Peter Kovesi @ CSSE UWA in Perth, WA, who made me know the Scale Space Theory and provided me with the Fast Radial news, paper and Matlab code. Cheers Peter.

Many thanks to Prof. Bellomi and Dr. Rampinelli of IEO, who provided the databases employed in this work.

Per il resto, sono fatti personali.