

AD-A150 940

DESIGN TUNING AND PERFORMANCE EVALUATION OF AN
AUTOMATED PULMONARY NODULE. (U) ROCHESTER UNIV NY DEPT
OF COMPUTER SCIENCE W LAMPETER FEB 83 TR-120

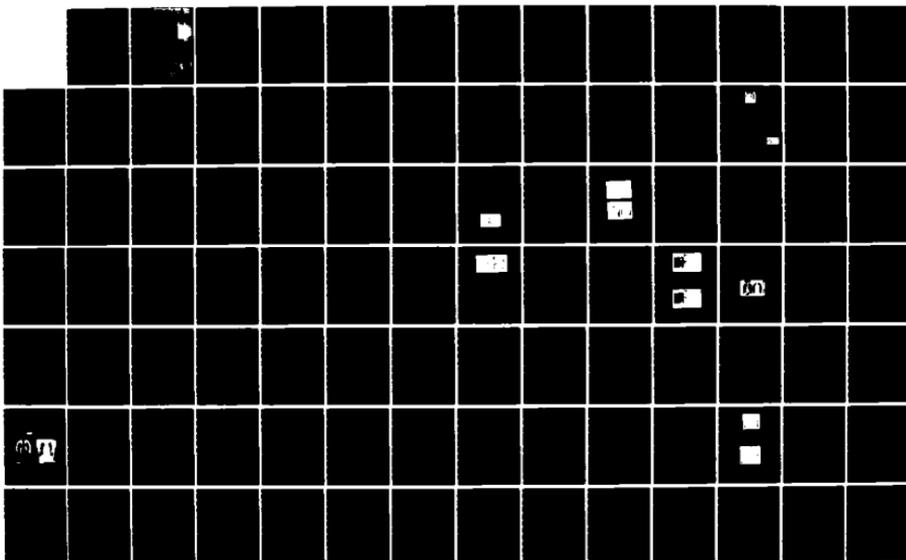
1/1

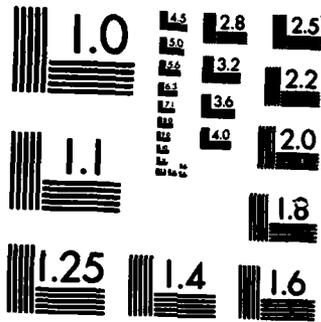
UNCLASSIFIED

N00014-80-C-0197

F/G 6/5

NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD-A150 940

DESIGN, TUNING AND PERFORMANCE
EVALUATION OF AN AUTOMATED
PULMONARY NODULE DETECTION SYSTEM

by

William Lampeter

TR 120

February 1983

DTIC FILE COPY

Rochester

Department of Computer Science
University of Rochester
Rochester, New York 14627

DTIC
ELECTE
MAR 04 1985

E

This document has been approved
for public release and sale (in
unlimited quantities) by the
DTIC Office

85 02 20 017

DESIGN, TUNING AND PERFORMANCE
EVALUATION OF AN AUTOMATED
PULMONARY NODULE DETECTION SYSTEM

by

William Lampeter

TR 120

February 1983

DTIC
COLLECTED
MAR 04 1985
E

This report reproduces a thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the School of Photographic Arts and Sciences in the College of Graphic Arts and Photography of the Rochester Institute of Technology

I wish to thank the Computer Science Department of the University of Rochester for the kind use of their facilities and for the honor of having access to their faculty. All the research contained in this thesis and the preparation of this document were completed at the University of Rochester.

This work was funded in part by the following: ONR (Office of Naval Research) grant N00014-80-C-0197; DARPA (Defense Advanced Research Projects Administration) grant N00014-82-K-0193; and NIH (National Institute of Health) Public Health Service grant 5R01-HL21253-05.

This document has been approved
for release and sale to
the public in its entirety.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM	
1. REPORT NUMBER TR 120	2. GOVT ACCESSION NO. AD A150 940	3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) Design, Tuning and Performance Evaluation of an Automated Pulmonary Nodule Detection System		5. TYPE OF REPORT & PERIOD COVERED Technical Report	
		6. PERFORMING ORG. REPORT NUMBER	
7. AUTHOR(s) William Lampeter		8. CONTRACT OR GRANT NUMBER(s) N00014-80-C-0197 and N00014-82-K-0193	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Computer Science Dept., University of Rochester, Rochester, NY 14627		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS	
11. CONTROLLING OFFICE NAME AND ADDRESS Defense Advanced Research Projects Agency 1400 Wilson Boulevard Arlington, VA 22209		12. REPORT DATE February 1983	
		13. NUMBER OF PAGES 102	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Office of Naval Research Information Systems Arlington, VA 22217		15. SECURITY CLASS. (of this report) Unclassified	
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE	
16. DISTRIBUTION STATEMENT (of this Report) Distribution of this document is unlimited.			
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)			
18. SUPPLEMENTARY NOTES None.			
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)			
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Radiologists miss approximately 25-30% of all pulmonary nodules smaller than 1.0 cm. in mass screenings. This paper describes a system for the automated detection of pulmonary nodules. It aids the radiologist by indicating the sites in the radiograph most likely to be nodules. Procedurally-driven image experts that respond to specific types of anatomic features are incorporated in a pattern recognizer which uses linear discriminant			

analysis to classify the candidate nodule sites. Sites not classified as nodules are eliminated from the list of sites presented to the radiologist for inspection. This system has been tested on 43 chest radiographs, and has demonstrated that pattern recognition techniques and procedurally-driven image experts are capable of reducing the number of sites that a radiologist must inspect from at most 17 to at most 3 in order to be 99% confident of having inspected any nodule detected by the system that is trained with 37 films.

ABSTRACT

Radiologists miss about 25-30% of all pulmonary nodules smaller than 1.0 cm. in mass screenings. A system for the automated detection of the pulmonary nodule has been designed, tuned, and tested on a 43 chest radiographs. The goal of this system is to aid the radiologist in locating a pulmonary nodule by indicating a few sites in the radiograph that are most likely to be nodules.

Procedurally driven image experts that respond to specific types of anatomic features have been devised and are incorporated in a pattern recognizer, which uses linear discriminant analysis, to classify the candidate nodule sites. Candidate nodule sites that are not classified as nodules are eliminated from the list of sites that are presented to the radiologist for inspection.

This work has demonstrated that pattern recognition techniques and procedurally driven image experts are capable of reducing the number of candidate nodule sites that a radiologist must inspect from at most 17 to at most 3 in order to be 99% confident of having inspected any nodule detected by the system that is trained with 37 films. The radiologist must be willing to accept a film true positive rate of 88% (as opposed to a film true positive rate of 92%) for the convenience of having fewer points to inspect.

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification _____	
By _____	
Distribution/ _____	
Availability Codes	
	Avail and/or
Dist	Special
A-1	



TABLE OF CONTENTS

1 - INTRODUCTION - Statement of Problem	1
1.1 Occurrence of the Solitary Pulmonary Nodule	3
1.2 Appearance of the Pulmonary Nodule	5
1.3 Detection of the Pulmonary Nodule by the Human Viewer	9
1.4 Previous Work: Digital Processing and Analysis of Chest Radiographs	13
2 - INTRODUCTION - Introduction to ANDS	17
2.1 Photographic Reduction and Digitization: Creation of a Test Database	21
2.2 Preprocessing: Spline Filtering and Histogram Equalization	25
2.3 Candidate Nodule Detection	31
2.4 Performance Evaluation	35
2.5 Incorporation of AI to Reduce the Number of False Positives	41
3 - EXPERIMENTAL	57
3.1 Optimal Reproduction of ANDS Database	59
3.2 Choosing an Image Processing Configuration	65
3.3 Tuning the Parameters of ANDS	67
4 - RESULTS - Evaluation of the Performance of AI in Reducing the False Positive Rate	73
5 - DISCUSSION	87
6 - CONCLUSION	89
7 - FUTURE WORK	91
8 - REFERENCES	93
APPENDIX 9.1 - Calibration of Optronics Scanner	97
APPENDIX 9.2 - Computation of ANOVA: Processing Methods	99
APPENDIX 9.3 - Description of ANDS Database and Summary of Nodule Detection Results	101

1 - Statement of Problem

The fundamental goal of this work is to improve the detection of the pulmonary nodule in chest radiographs. Despite improvements in radiographic imaging technology, radiologists are unable to detect approximately 30% of all pulmonary nodules smaller than 1.0 cm in mass screenings. The most dangerous type of pulmonary nodule, the malignant lesion, is most difficult to detect in its earliest stages. If the radiologist were able to detect such lesions in their early stages the patient would often have a better prognosis for survival and the consequent treatment would be less radical. Besides easing the radiologists' workload, automated nodule detection would provide a means of recording sites of possible nodules that should be monitored in subsequent films and a means for evaluating the performance of radiographic imaging processes.

It is known that the human viewer has difficulty detecting small lesions in chest films. That is, given a film that contains a small (0.5 cm. or less) lesion the human viewer will often fail to detect it. However, if the lesion were pointed-out it would be recognized. It is believed that this inability is due to limitations of the human visual system in detecting objects against a background of (structured) noise. The computer is immune to influences from structured noise. It can tirelessly search an entire radiograph and report the presence of all small round things. The ANDS (Automated Nodule Detection System) processes a chest radiograph and provides a display of sixteen or fewer sites in the chest radiograph that are most likely to be a nodule.

The possible benefit of ANDS is evident when one considers that about 30% of all *visible* solitary pulmonary nodules go undetected in routine viewing of chest radiographs [Garland, 1959][Yerushalmy, 1951]. This limitation is presumably not due to radiographic technology but is inherent in the human observer. The human observer can reliably diagnose pulmonary nodule's 1.0 cm. or larger in diameter but exhibits decreasing proficiency as the nodule diameter gets smaller. The radiograph is capable of representing a nodule as small as 0.3 cm. which must often go undetected for nine or more months until it reaches sufficient size to be seen [Goldmeier, 1965]. Since roentgen findings are usually present in presymptomatic stages, their recognition at early stages is presumably of incalculable benefit to the patient. Automated diagnosis of the pulmonary nodule offers the hope of easing the radiologists workload by helping to limit the search area.

The specific goal of this work is to improve an automated nodule detection system that guides a human viewer to sites on a chest film that are most likely to be pulmonary nodules and to reduce the number of false positives that are reported by that system. The following processes comprise the current ANDS: 1) photographic copying and digitization of radiograph; 2) image processing to enhance the appearance of the nodules; 3) candidate nodule detection and accumulation of votes; and 4) elimination of false positives. The main features of the ANDS are: spline filtering to subtract out background variation; candidate nodule, rib, and vascularity detection using Hough-like techniques; and discriminant analysis to reduce the rate of false positives. These are further discussed in Chapter 2.5. Note that *false positive* as it is used with respect to the ANDS has the following meaning: any non-nodule that is considered *more nodule like* than a nodule is a false positive. This concept will be clarified in the section on nodule detection and accumulation of votes. The body of the experimental work was to design and tune the ANDS. This work was done in four phases: optimization of photographic/digital reproduction of the radiograph; evaluation of four nodule detection processes; parameter tuning of the chosen process; and discrimination of false positives. These are discussed in the Experimental section. The following section is an introduction of the problem of nodule detection. It is a survey of image processing of chest radiographs and an overview of the work done in the computer analysis of chest films. It serves as both a justification and as a motivation for this work.

1.1 - Occurrence of the Solitary Pulmonary Nodule

The prevalence of the solitary pulmonary nodule (SPN) has been reported between 1 per 1000 (films studied) [Holin, 1959] and 2 per 1000 [Good, 1958]. The study by Holin, a community wide survey in Cincinnati (1949), consisted of 673,281 films with 687 diagnosed as containing nodules. Over the last four decades the incidence of bronchogenic carcinoma has quadrupled in industrialized countries [WHO, 1965]. Deaths due to bronchogenic carcinoma were: in Japan, 1.3 per 1,000,000 in 1950 and 6.5 per 100,000 in 1960; in Great Britain, 50 per 100,000; and in the U.S. 20 per 100,000 in 1960 [WHO, 1965]. The likelihood that a nodule is malignant increases with the age of the patient. Walske reports malignancy in 53% of all cases aged 50 or more years and 12% malignancy in patients under age 50 [Walske, 1966]. Steele found malignancy in 56% of all cases aged 50 or more years [Steele, 1963]. In a review of 25 case studies which involved a total of 1203 patients, malignant lesions were found on the average to comprise 36.7% of all lesions; the percentage of malignant nodules varied in the case studies between 7% (Jones & Cleave, 1954) and 78% [Axtmayer & Ehrlich, 1955] [Davis, 1956]. Seybold reports an average malignancy of 37.8% in his survey of 22 case studies which involved a total of 2258 cases; the data ranged between 7% (1 of 14 cases - Jones & Cleave, 1954) and 55% (37 of 67 cases - Husfeldt & Carlsen, 1950) [Seybold, 1964]. It should be noted that the criteria for inclusion in each of these 25 studies varied as did the sex and average age of the patients included. Some studies took place in veterans hospitals which were predominated by older males while others took place in armed services hospitals which are mostly comprised of younger males. Lung cancer has been shown to be about four times more prevalent in males (21% of all males cancers) than in females (5% of all female cancers) [AMN, 1973].

The prevalence of the solitary pulmonary nodule (SPN) has been reported between 1 per 1000 (films studied) [Holin, 1959] and 2 per 1000 [Good, 1958]. The study by Holin, a community-wide survey in Cincinnati (1949), consisted of 673,281 films with 687 diagnosed as containing nodules. Over the last four decades the incidence of bronchogenic carcinoma has quadrupled in industrialized countries [WHO, 1965]. Deaths due to bronchogenic carcinoma were: in Japan, 1.3 per 1,000,000 in 1950 and 6.5 per 100,000 in 1960; in Great Britain, 50 per 100,000; and in the U.S. 20 per 100,000 in 1960 [WHO, 1965]. The likelihood that a nodule is malignant increases with the

age of the patient. Walske reports malignancy in 53% of all cases aged 50 or more years and 12% malignancy in patients under age 50 [Walske, 1966]. Steele found malignancy in 56% of all cases aged 50 or more years [Steele, 1963]. In a review of 25 case studies which involved a total of 1263 patients, malignant lesions were found on the average to comprise 36.7% of all lesions; the percentage of malignant nodules varied in the case studies between 7% (Jones & Cleave, 1954) and 78% [Axtmayer & Ehrlich, 1955] [Davis, 1956]. Seybold reports an average malignancy of 37.8% in his survey of 22 case studies which involved a total of 2258 cases; the data ranged between 7% (1 of 14 cases - Jones & Cleave, 1954) and 55% (37 of 67 cases - Husfeldt & Carlsen, 1950) [Seybold, 1964]. It should be noted that the criteria for inclusion in each of these 25 studies varied as did the sex and average age of the patients included. Some studies took place in veterans hospitals which were predominated by older males while others took place in armed services hospitals which are mostly comprised of younger males. Lung cancer has been shown to be about four times more prevalent in males (21% of all males cancers) than in females (5% of all female cancers) [AMN, 1973].

1.2 - Appearance of the Pulmonary Nodule

A solitary nodule is a circumscribed mass situated in the substance of the lung, constituting the only significant pathologic process in the lungs of the patient being examined, and showing no significant signs of cavitation [hollowness] or obstruction of the airway [Good, 1963].

The SPN is also called a "coin lesion" because of its circumscribed circular shape. This is, however, a misnomer because the shape of the nodule in fact spherical or ovoid and not flat and round.

The SPN manifests itself in essentially two forms, as benign and malignant lesions. There are only two conditions under which a nodule may be considered benign: the presence of dense calcification or signs of stability lasting two or more years [Good, 1963].

Density, shape, size, location, and margination (border characteristics) are considered as possible measures to aid the discrimination of nodules from other objects within the lungs. It is important to note that many of the following statistics refer to nodules which were most likely diagnosed because they were *seen* in chest radiographs and/or because they were resected (surgically removed). The following statistics which regard the SPN may not necessarily apply to the general population of radiographic images of all SPNs, especially small barely perceptible ones. Many of the following statistics describe that population of visible nodules which are 1.0 cm or larger and whose benign/malignant nature is known.

density

Perhaps the most powerful discriminating feature of the SPN is optical density. Dense images of nodules often indicate the presence of calcification - the primary feature of the benign nodule. The density of the nodule in the film may be used to distinguish the benign from the malignant nodule. However, the presence of some calcium does not indicate that a nodule is benign. Ten cases of 280, 3.7 percent, of primary carcinoma were read as containing some calcium [Steele, 1963]. Dense nodules are less likely to be considered malignant [Vivas, 1953] [Steele, 1963] [Davis, 1956]. Small

lesions of heavy density are less likely to be malignant [Vivas, 1953]. Lesions smaller than 2.0 cm and dense are usually granulomas (benign); lesions larger than 2.0 cm in diameter and not dense are probably carcinomas [Davis, 1956]. Malignancy is unlikely in a dense or concentrically calcified nodule [Steele, 1963]. Siegelman uses density as a means of discriminating between calcified-benign and malignant nodules in CT images. He found that no malignant lesion in his study had a CT number greater than 147 Hounsfield units [Siegelman, 1980]. His CT system was carefully calibrated to produce quantitative results.

shape

A majority of carcinomas are characterized by an irregular shape [Good, 1953]. [Steele, 1963]. Thirty-seven percent of the carcinomas studied were irregular or elongated and only 11% had sharp margins [Steele, 1963]. Shape offers little discrimination between benign and malignant nodules.

size

Although nodules approximately 3 mm in diameter are visible, the lower limit for diagnosis is believed to be 1.0 cm [Goldmeier, 1965]. In a study of 1267 nodules, 714 could be measured either radiographically or pathologically; 66% were 5 cm or greater, and 1.26% were less than 1.0 cm in diameter [Theros, 1977]. A greater proportion of nodules was found to be malignant as the nodule diameter increased [Holin, 1959]. Davis reports in his case study that the smaller nodules are most likely to be considered benign [Davis, 1956]. One might infer from the following that small malignant nodules are often not found. In a review of 22 case studies, Seybold concludes that few lesions greater than 5 cm were benign and few less than 2 cm were malignant [Seybold, 1964]. Holin reports that the average size of malignant nodules in his case study as 5.2 cm, and 2.5 cm as the average size for tuberculosis nodules [Holin, 1959]. As the size of the bronchogenic carcinoma shadow increases: operating becomes more difficult; post-operative mortality is higher; and the overall prognosis is worse [Bateson, 1964].

Goldmeier hypothesizes that small nodules may have a lower visibility because nodules have a 2.5

1.2 - Appearance of the Pulmonary Nodule

mm outer shell of low density. In order to be visible the nodules must be considerably larger than 5.0 mm [Goldmeier, 1965].

location

Tuberculous granulomas were found predominantly in upper lobes [Bateson, 1965]; and three times more frequently in the upper lobe than in the lower [Steele, 1963]; solitary metastases were found predominantly in the lower lobes, otherwise no other lesions showed particular distribution [Bateson, 1965]. Davis found, in his study of 215 cases, that the distribution of nodules in the lobes was fairly even with no great distribution differences between benign and malignant nodules [Davis, 1965]. Holin in his community-wide study reports that 61% of the lesions found were in the right lung and 39% were in the left lung; more nodules were found in the lateral (68%) than in the medial portions of the lung (32%) [Holin, 1959]. The above results may either represent reality or serve as an indictment of the proficiency of the human diagnostician.

margination

Most radiologists argue that it is impossible to differentiate benign from malignant nodules on the basis of size or margination... [Davis, 1956]. Of the 100 solitary circumscribed carcinomas included in Bateson's study, the shadows of 71% had an ill-defined margin. He reports a tendency for a higher proportion of shadows of small carcinomas to be ill-defined and the shadows of large carcinomas to be well-defined. He also reports that the prognosis is better for patients with nodules with well-defined shadows [Bateson, 1964]. A ragged, fuzzy edge and an irregular outline are more often present in primary cancer [primary as opposed to a metastatic, or secondary manifestation of cancer][Seybold, 1964]. In general, carcinomas tend to be less well-defined; the sharpness of the border could not be used to distinguish between benign and malignant [Good, 1963].

In conclusion, the most harmless nodule is perhaps the easiest to detect in its early stages; and the nodule which presents the greatest danger to the patient is seemingly the most difficult to detect in its early stages. For example, a possible interpretation of Davis' finding, that small nodules are

most likely to be considered benign, is that malignant nodules are simply not likely to be discovered until they are big. A similar interpretation may also be applied to Seybold's finding that few nodules less than 2 cm were found to be malignant [Seybold, 1956] - perhaps few small malignant nodules could be seen. Malignant nodules which are claimed to have irregular shapes and ill-defined margins [Seybold, 1964] [Good, 1963] are presumably more difficult to detect because these features are presumably more difficult to detect. The statistics which point to a high incidence of malignancy among large nodules should serve as an indication of the necessity to find a means of detecting malignant nodules when they are small.

Since the goal of the nodule detector is to find any nodule, no distinction (that is, based on brightness) is made between benign and malignant nodules. The centers of candidate nodules (sites in the chest film that are most likely to be nodules) are detected by a Hough like circle transform. The Hough circle transform has been generalized to permit detection of a variety of bright closed shapes. Thus the circle transform is able to detect a variety of closed shapes. Nodule appearance characteristics that are important to human detection of the pulmonary nodule have been incorporated in ANDS. Knowledge about the relative brightness and irregular shape of the pulmonary nodule is embedded in the CN Expert (a program that uses procedural knowledge to locate the center of a CN). Appearance characteristics of the nodule border: contrast, and azimuthal uniformity are used to discriminate nodules from the non-nodules among the CNs that are reported by ANDS. Global knowledge, knowledge about the nodule in relation to its environment, is also used to discriminate nodules from non-nodules.

1.3 - Detection of the Pulmonary Nodule by the Human Viewer

In addition to the intrinsic features mentioned in section 1.2, the surrounding anatomical contrast greatly effects the visibility of a nodule. Nodules of decreasing contrast are increasingly difficult to detect [Kundel, 1979]. Nodules with sharper edges are identified faster and with greater frequency than those with less sharp edges [Carmody, 1980]. Nodules of decreasing size are increasingly difficult to detect. Detection accuracy rates of 44% for 1.0 to 1.5 cm, and 8% for 0.5 cm nodules are reported [Kelsey, 1977]. The effect of the surround complexity, i.e. anatomical *busyness*, is suggested in the finding, that when the same nodule was superimposed in various lung regions 56% of upper-left and 29% of lower-left lesions were seen [Kelsey, 1977]. Note the coincidence of this finding with that of Steele - that tuberculous granulomas are found three times more frequently in the upper than in the lower lobes. Kruger *et al.*, who automated the classification of coal workers' pneumoconiosis, report that their device correctly classified 77% of the disease in the lower left and 81.5% in the upper left lung. This disparity between human and automated detection in these lung regions suggests that a perceptual rather than a pathological basis may be responsible for these findings and that automated methods may exhibit less error.

Viewing distance and brightness level have been shown to effect nodule detection. Shea *et al.* have found that the peak of the VSTF (Visual System Transfer Function) decreases in amplitude and frequency as brightness decreases. They propose that every abnormality has a unique (optimal) viewing distance [Shea, 1977]. Hemmingsson *et al.* superimposed 2.0 cm diameter lesions in chest films and found that a density difference of 0.025 to 0.060 between the nodule and surround was necessary for detection. They also found that the density difference at which a nodule is first discernable is a function of viewing distance and the characteristics of the object's border and adjacent structures. They suggest that the optimum viewing distance varies for different lesions [Hemmingsson, 1975].

The parameters of the conspicuity metric, K_2 , which was determined by Revesz *et al.* are those features which distinguish *seen* and *unseen* nodules. The features of an *undetected-but-visible* nodule were compared with those from the same nodule which was only detected in a later film. This metric was found to best distinguish between populations of detected and undetected nodules. This

metric has as its parameters: edge gradient and contour (steepness and smoothness, respectively) which are represented in the parameter EI, Edge Index; and surround complexity. These parameters provided better discrimination between the two populations of nodules than any of the following parameters tested: size, density difference between nodule and surround, and the rate of change of density around the nodule border as determined by the Laplacian [Revesz, 1977]. Edge gradient and edge uniformity are used by ANDS to discriminate among CNs. These are incorporated in ANDS as Edge Strength and Edge Visibility, respectively. The standard deviations of these measures are in fact used in conjunction with other measures by the pattern recognizer to classify CNs.

Errors in detection of nodules may occur in four levels of the search process: orientation, search, recognition, and decision making [Kundel, 1978]. Orientation errors occur when the observer is unfamiliar with chest films and cannot differentiate abnormal objects from background features. A search error occurs when an area containing a nodule is overlooked. When a nodule is scanned over but not recognized, a recognition error occurs. Decision making errors occur when an ambiguous figure is recognized but either falsely accepted or rejected. Kundel *et al.* claim that scanning errors account for 30% of all detection errors; recognition errors 25%; and decision-making errors 45% among skilled observers [Kundel, 1978].

Studies of film reader error over the past twenty years indicate error rates (films missed) of 25-30%. Despite advances in radiograph technology, readers have not been able to find more nodules. Reading errors may be attributed to faulty processing of visual information which falls into two domains: perceptual (unconscious process), and cognitive (conscious process).

Spatial vision research perhaps offers some explanations for human limitations in nodule detection. The importance of foveal vision in nodule detection suggests that high spatial frequencies influence detection. Where, "...small objects like pulmonary nodules can only be perceived if they are close to the center of the visual field. The more complex the visual information, the closer to the center of the visual field (e.g. the fovea) a small object must be imaged to be perceived [Kundel, 1978]. In order for a square wave to be distinguished from its fundamental sinusoid, at least its third harmonic must exceed visibility threshold [Campbell, 1968]. Similarly, an edge will go undetected if its high

1.3 - Detection of the Pulmonary Nodule by the Human Viewer

frequency components are below contrast threshold.

The effects of structured noise on the surround have a significant effect in how humans detect nodular abnormalities [Kundel, 1975]. Structured noise decreases the possibility of detection and increases the time for a reader to make a response [Kundel, 1975]. The effects of surround may be generalized to fall into two categories: overlapping and non-overlapping [Kundel, 1978]. The non-overlapping surround contributes to the overall complexity of the image, acting as a camouflage, exerting its effect on visual search rather than on visibility [Kundel, 1978]. An overlapping surround leads to edge obliteration and causes difficulties in both detection as well as in physical measurements [Kundel, 1978]. An occluding rib is an example of such a surround effect. Foveal performance is inhibited by the presence of extra stimuli in the periphery as well as in the fovea itself [Mackworth, 1965]. Diminution in the ability to perceive a given spatial frequency, e.g. the third or higher harmonics which would characterize the edge, may be due to lateral inhibition from the surround. Adaptation of cats to a given spatial frequency has been shown to raise the contrast required to produce a given response by a factor of about four [Movshon, 1979]. The effect of structured noise is evident in the failure of image processing, "TV processing" according to Kundel, to make nodules any more visible [Kundel, 1968]. Extensive processing makes nodules more conspicuous if their locations are known in advance, while detection is more difficult if their locations are unknown [Kundel, 1975]. This may represent a structured noise effect since processing may increase the structured noise more than it enhances the target abnormality [Kundel, 1975].

TR - 120

1.4 - Previous Work: Digital Processing and Analysis of Chest Radiographs

There are two main aspects to automated diagnosis of radiographs, the diagnosis of pattern, and the acquisition and analysis of large amounts of data [Henschke, 1979]. The acquisition and analysis of data preceded the analysis of image patterns, with the former occurring in the early to-mid 60's and the latter in the early 70's. Diagnosis of patterns in the chest encompasses measurements for the diagnosis of rheumatic heart disease [Hall *et al.*, 1971]; classification of coal workers' pneumoconiosis [Hall *et al.*, 1976] [Jagoe *et al.*, 1975]; analysis of pulmonary infiltration [Tully, 1978]; and detection of the solitary pulmonary nodule [Ballard, 1973]. The acquisition and analysis of large amounts of radiographic data, symptoms, and test results to arrive at a diagnosis has as its main stumbling block the inconsistent interpretation of the data from the radiograph by a human observer into a form amenable to processing by a computer [Henschke, 1979].

Coding radiographs is perhaps the first recorded instance of automated diagnosis. This involves quantifying aspects of the visual image into numerical sequences which are amenable to computer analysis. In this method the radiologist codes observations for computer analysis [Lodwick, 1963]. It was found that the problem inherent in handling such data is the conversion of the visual data into the exact qualitative and quantitative forms required by the computer [Lodwick, 1963]. Yamamura *et al.* point out similar difficulties in coding radiographic findings as *inconvenient in reproducibility and different readings from different viewers* [Yamamura, 1965]. They conclude that *...the highly complicated findings of the pulmonary lesion are beyond the ability of pattern recognition of an electronic computer. They are [best] left to the management of human brains* [Yamamura, 1965]. Meyers *et al.* digitized a radiograph using a flying spot scanner and displayed the image on an oscilloscope; they also displayed the derivative function of the image. They report that the radiographic image retrieved from their computer is *"...the most informative image of a portion of the lung and ribs that [they] have ever seen."* Furthermore, to the encouragement of computer visionaries, they predicted that digital analysis of mechanically scanned radiographs would be possible [Meyers, 1963].

Kundel *et al.* suggest that digital image processing of the chest radiograph is necessary to reduce the near 30% false negative of the human viewer; they outline image processing techniques [Kundel,

1969]. Moore claims that image processing would probably make a useful contribution to radiology by clarifying pictures of low quality [Moore, 1969]. Moore predicted that image processing of radiographs would permit:

- removal of image noise
- correction for geometric distortion
- elimination of non-uniform brightness,

and automated analysis techniques would:

- search for tuberculosis and heart enlargement in mass screenings
- determine physiological age from xrays of the hands
- detect lesions in mammograms
- predict time of tooth eruption
- analyze angiograms
- calculate bone densities.

Fourier filtering techniques to enhance the appearance of the pulmonary nodule were evaluated by Ziskin. He found these techniques incapable of separating the nodule from its surround [Ziskin, 1972]. Similarly, Kundel found that processing the radiographic image did not lead to an increase in nodule detection [Kundel, 1975].

The analysis of pulmonary infiltrates and classification of pneumoconiosis is essentially a problem of texture analysis. Quantitative texture measures are used to distinguish between normal lung, alveolar infiltrates, and interstitial infiltrates with 95% accuracy in the training phase, and 90% accuracy in the testing phase [Tully, 1978]. The image texture is analyzed using Ausherman's SGLDM (spatial grey level dependence method). This method is essentially statistical; its measures are based on the probability of going from a specific pixel value to another specific value at a given point in a textured image. Differences in image quality due to exposure time and development conditions are eliminated by linearly redistributing the image so that contrast is normalized and the number of grey levels is reduced - less grey levels lead to greater accuracy when using the SGLDM [Tully, 1978].

1.4 - Previous Work: Digital Processing and Analysis of Chest Radiographs

Two systems for the detection and classification of coal workers' pneumoconiosis have been described, one in the U.S. by Kruger *et al.* and one in Great Britain by Jagoe and Paton. Federal laws enacted in 1969 require that coal workers be regularly examined for pneumoconiosis. Criteria for classifying the severity of the disease have been standardized. The U.S. system employs opto-digital analysis while the U.K. system operates entirely on digital images. The opto digital method involves imaging the Fourier spectrum of the radiograph with a laser and then analyzing the spectrum with annular wedges to extract a frequency signature which is then subjected to statistical classification using linear discriminant functions [Kruger, 1977]. Jagoe and Paton's method for classifying pneumoconiosis involves measuring the unevenness of the density distribution within square grids 3.6mm on a side. The diagnoses by their process have demonstrated a 0.88 correlation with those by radiologists [Jagoe, 1975].

The earliest known work on the automated detection of the solitary nodule may be attributed to D.H. Ballard and J. Sklansky [Ballard, 1973]. This work involves image processing to enhance detection of a tumor edge in digital representations of chest radiographs and radioisotope liver-scans. The detection of this edge was deemed difficult for two reasons: changes in the image density about the perimeter of the nodule which are caused by background density gradients; and the presence of ribs which may occlude the nodule [Ballard, 1973]. The work done by Ballard is the foundation of the ANDS and this thesis.

TR - 120

2 - Introduction to ANDS

A chest radiograph is the input of ANDS which processes and analyzes the image for CNs (Candidate Nodules). The output of ANDS is a list of CN sites which are displayed for review by a human viewer. Fig. 2.0.1 presents the essence of ANDS. Table 2.0.1 illustrates the steps in ANDS, their inputs, outputs, and effects. The purpose of the first step, photographic reduction and digitization of the chest radiograph is to render the 14"x17" chest film into a form amenable to digital processing. The photographic reduction step is necessary because the available image digitizer is not capable of digitizing any image that is larger than 10"x10". The goal of this stage is to achieve a linear mapping between optical densities (in the lung parenchyma, that is, in the lung tissue) and pixel values and to maintain the required spatial resolution. In a pre-processing step the background variation is removed using a spline filter and the contrast is enhanced with histogram equalization. CNs are located using a Hough-like technique, which votes for CNs in an accumulator array whose dimensions correspond to the image dimensions. The peaks in the accumulator array correspond to the locations of centers of closed circular shapes. Following the application of the Hough technique the accumulator array is smoothed by convolution with a Gaussian operator. This improves the estimate of the center of a CN, which is represented as a local peak. The smoothed accumulator array is searched for a specified number of the highest valued peaks. The locations of these peaks correspond to the locations of centers of CNs. The locations of the nodules in the films that were tested are known. A metric has been devised to measure the performance the nodule detection process. This metric uses the list of CN locations that is produced by ANDS and the locations of the known nodules. Since some of the reported CNs are obvious errors, for example, lung borders and ribs are common false positives two procedurally driven recognition experts and a technique for linear discriminant analysis have been incorporated in ANDS to reduce the false positive rate. Each of the stages in ANDS is described in greater detail in the following sections.

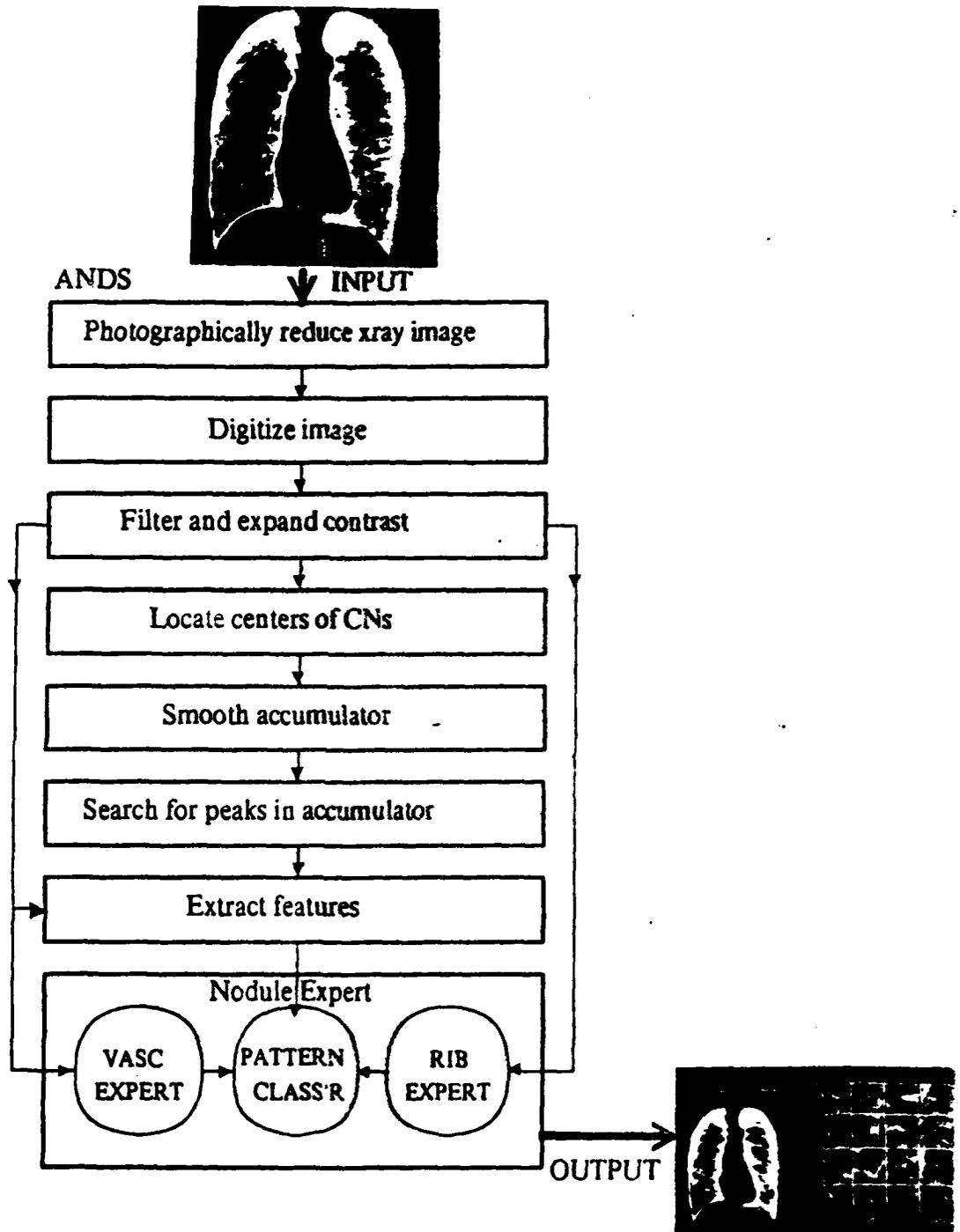


Figure 2.0.1 - The Automated Nodule Detection System - ANDS. A chest radiograph is presented to ANDS which analyzes it for the presence of pulmonary nodules and which produces a display of candidate nodule sites. ANDS incorporates three vision experts: a Rib Expert and a Vascularity provide information to a pattern classifier which classifies a CN; and a Nodule Expert which, using a set of rules, distinguishes nodules from false positives, causing the obvious false positives to be omitted from the CN sites that are presented to the radiologist.

2 - Introduction to ANDS

INPUT	PROCESS	OUTPUT
14"x17" chest film	Photographic reduction	4"x5" negative on Kodak Commercial film; size of radiograph is reduced to 0.26 times original
4"x5" negative	Digitization	digital image; 8-bits/pixel; sampled at 100µm; optical densities are converted to pixel values
digital image of chest radiograph	Spline filtering and histogram equal'n	enhanced digital image; background variation is removed and contrast is enhanced
spline filtered histogram equalized image	Candidate nodule detection	image that contains votes for locations of CN centers
accumulator image	Smooth accumulator image	smoothed image whose peaks represent the locations of CN centers; the groupings of votes cast in the previous step are concentrated about their center of mass
smoothed-accumulator image	Search smoothed accumulator image	list of CN center coordinates ordered by accumulator value;
accumulator list, spline filtered image	Elimination of false positives	modified ordered list of CN center coordinates; false positives reduced
file of known nodule locations, list of locations of CN centers	Performance evaluation	a report of the number of false positives, the true positive rate, the CIIM, and the DM

Table 2.0.1 - The inputs and outputs of ANDS.

TR - 120

2.1 - Photographic Reduction and Digitization: Creation of a Test Database

Anterior-posterior chest radiographs which are representative of the general population of such chest films were obtained from Dr. John Wandtke of the School of Medicine of the University of Rochester. The performance of ANDS was evaluated using these films. Hence, these films are referred to as the ANDS database.

Fifty 14"x17" chest radiographs, 44 containing at least one nodular abnormality and 6 normals, were photographically reduced and digitized. The dimensions of the digitized images are about 900x900 pixels. Fig. 2.1.1 illustrates the reproduction process. A Sinar "C" camera with a 240mm Xenar lens was used to image each radiograph onto Kodak Commercial film. The non-lung area of each radiograph was masked prior to copying using exposed xray film. Only the lung areas were imaged when copying the radiograph, i.e. no light was allowed to pass through the non lung area of the radiograph. This was done to reduce camera/lens flare in order to obtain a more linear transfer from optical density to pixel value, see Chapter 3.1. A 10" Kodak #2 step wedge and a tri-bar target were included when copying each radiograph. These provided means for quantitatively assessing the transfer of densities (tone reproduction), and assuring that a nominal (as given by the Nyquist sampling relation) spatial resolution was maintained. Since the radiograph was digitized on an Optronics C4100 rotating drum scanner at a contiguous sampling interval of 100 microns with circular apertures of 100 microns (illumination and collection), a spatial resolution in excess of the nominal 1.25 lp/mm was maintained. The film was developed in Kodak HC-110 developer, dilution D, for 5 minutes at $68 \pm \frac{1}{2}^{\circ}\text{F}$ with R.I.T. tray-rock agitation. The Optronics was calibrated, using a 5" Kodak #2 step wedge, to provide maximal useful range and optimal discrimination between densities around 2.65, the upper limit of the lung region densities in the photographic reduction. See Appendix 9.1 for details on the calibration of the Optronics scanner.

A statistical analysis was performed to determine the optimal exposure and flare condition. The exposure/flare condition that resulted in a statistically insignificant second-order term in a regression of pixel value as function of radiograph density was chosen from the 9 exposure/flare conditions on 3 representative films tested. This exposure/flare condition was used when the 50 films that constitute the database were copied. The nodule size/age, and patient-sex/disease/number-of-nodules-per-film distributions of the films in the database are illustrated in Figs. 2.1.2 and 2.1.3, respectively.

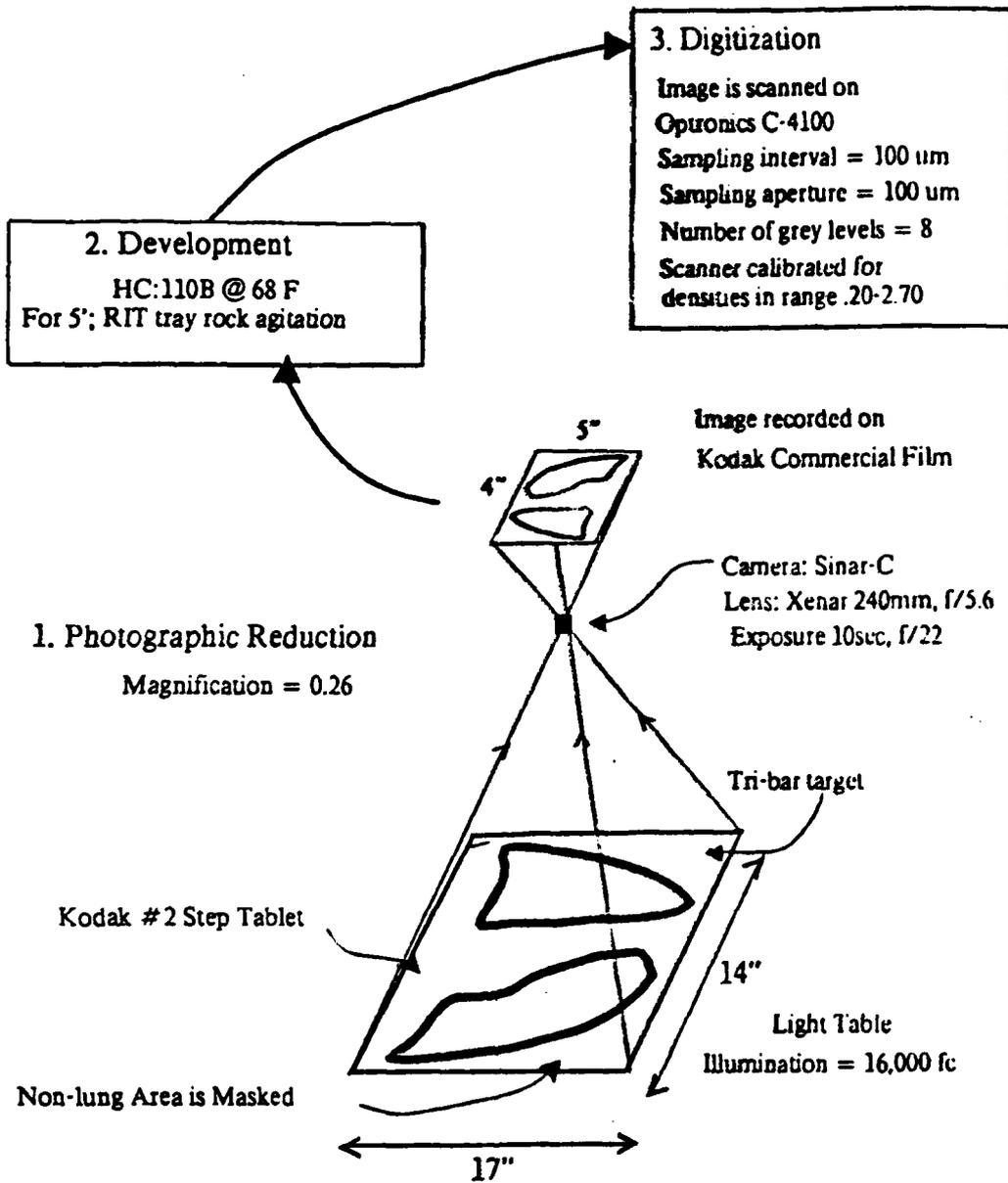


Figure 2.1.1 - The photographic reduction and digitization process that was used when copying the 50 chest radiographs that constitute the ANDS database. The 14"x17" radiograph is photographically reduced onto 4"x5" Kodak Commercial film by the camera/lens system. The exposed sheet film is developed. The developed film is digitized on an Optronics rotating drum scanner to an 900x900 pixel image.

Figure 2.1.1 - The photographic reduction and digitization process that was used when copying the 50 chest radiographs that constitute the ANDS database. The 14"x17" radiograph is photographically reduced onto 4"x5" Kodak Commercial film by the camera/lens system. The exposed sheet film is developed. The developed film is digitized on an Optronics rotating drum scanner to an 900x900 pixel image.

2.1 - Photographic Reduction and Digitization: Creation of a Test Database

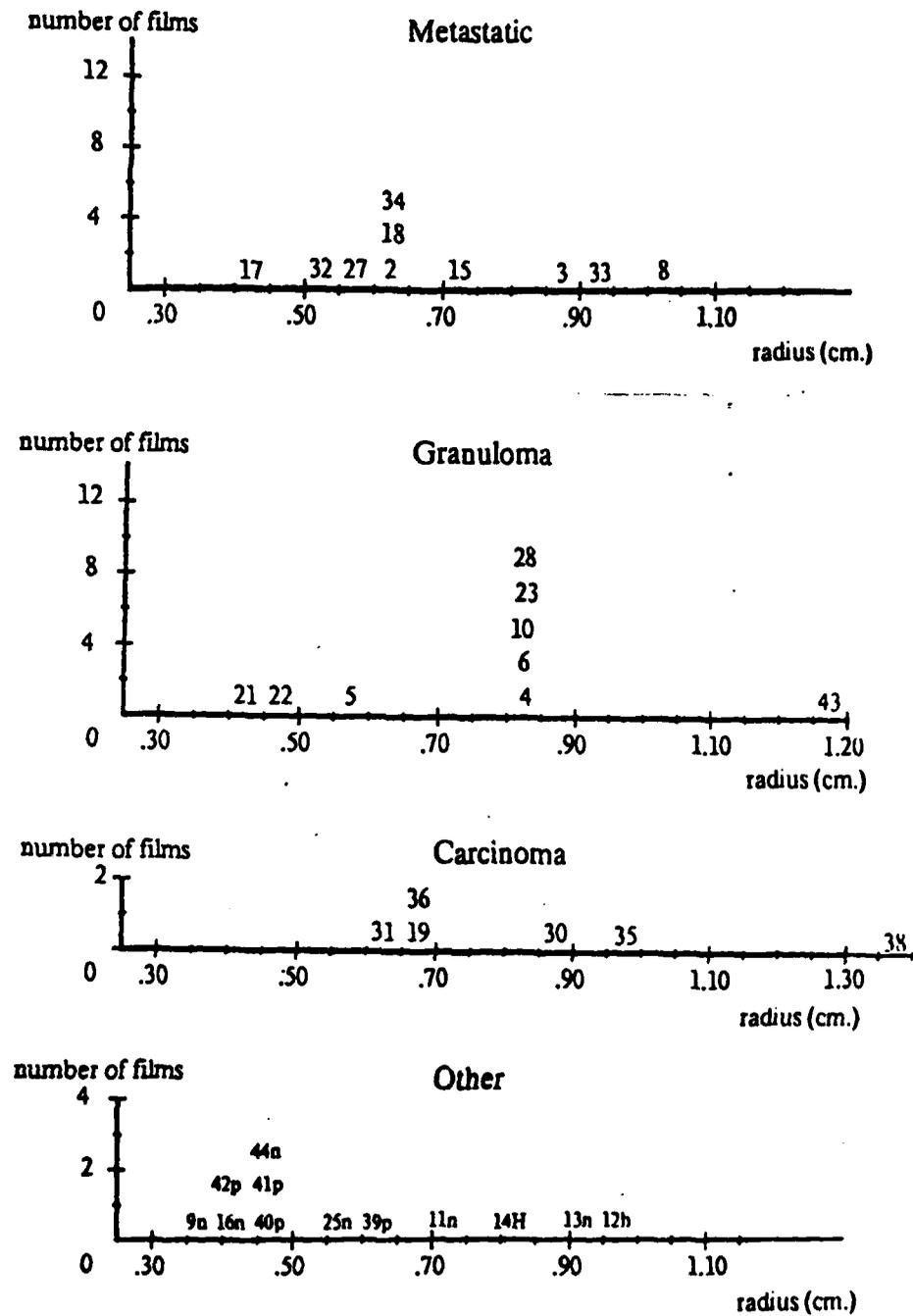


Figure 2.1.2 - Distribution of nodule radius by disease. The numbers are the identification numbers of the films that contain the nodule(s). The radius value is the average pixel boundary distance over all radial arms (see Fig. 2.5.5). The values are average nodule radius for films that contain more than one nodule. Other includes the following types of nodular abnormalities: nipple (n), hamartoma (h), Hodgkins disease (H), and pseudo-nodule (p).

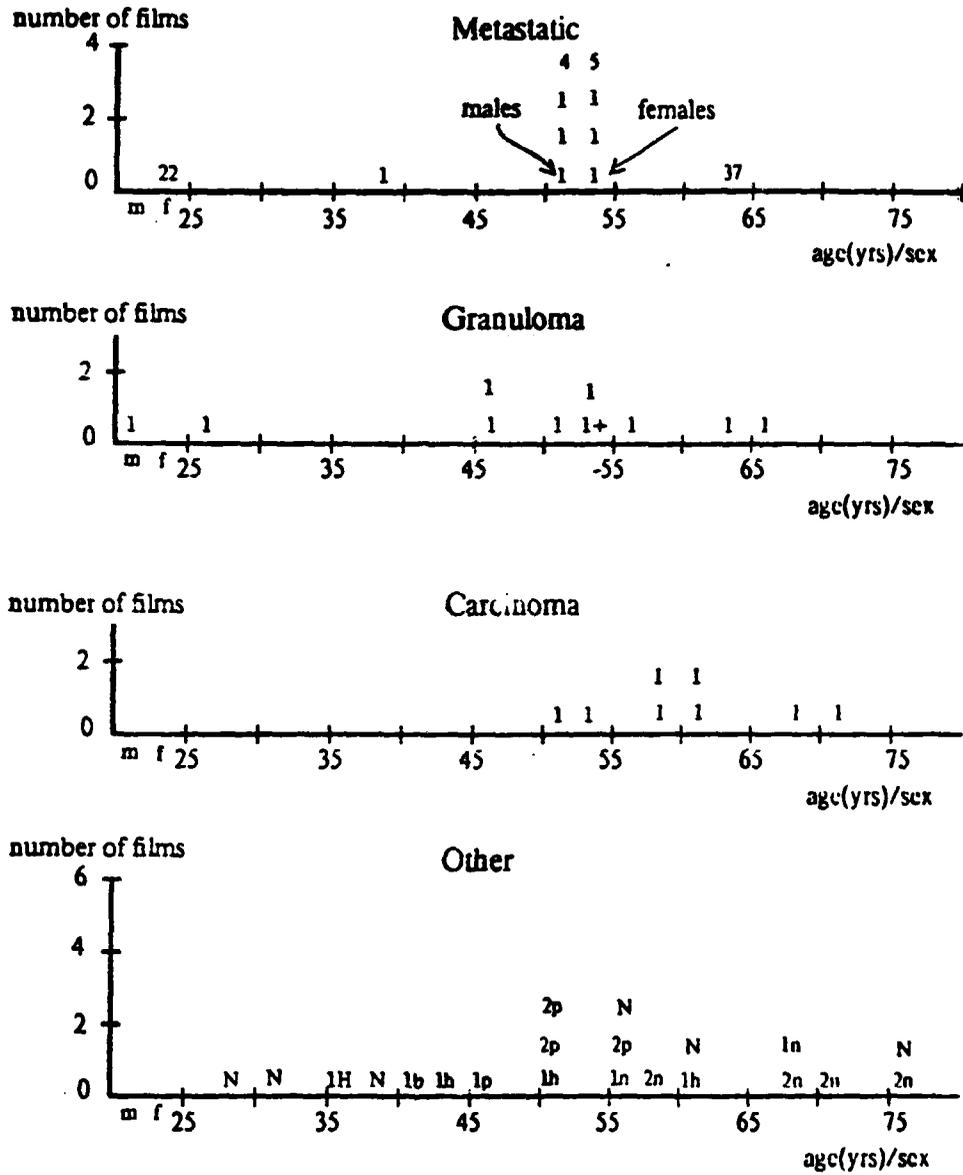


Figure 2.1.3 - Distribution of age, sex, and number of nodules per film by disease. Each age bracket has two fields which represent the sexes; male is the leftmost field and female is the rightmost field. The numbers that are plotted are the numbers of nodules per film. The letters that precede the numbers for pathology Other represent: N=normal, b=button, h=hamartoma, H=Hollands, p=pseudo nodule, and n=nipple.

2.2 - Preprocessing: Spline Filtering and Histogram Equalization

Spline filtering and histogram equalization serve to make the small details of the image more visible by subtracting background variation. Spline filtering is similar to field flattening [Pearson *et. al.*]. Essentially, the low frequency components of the image are removed when spline-filtering. A low frequency approximation of the image is made by interpolating with B-splines. This interpolated image is subtracted from the original image. The spline filter has three steps: interpolating the original image to produce a two-dimensional approximation; subtracting the interpolated image from the original image; and expanding the contrast of the spline filtered image using histogram equalization. The parameter of the spline filter is the interval at which the interpolant points are taken - the knot spacing. This interval corresponds to the number of points that are interpolated between knots. Figure 2.2.3 illustrates the effect of histogram equalization on an image that was filtered at two different knot spacings. As the distance between the sampled knots decreases the interpolated image more closely approximates the original image. More frequently sampled images contain more high frequency content; this is evident in Figure 2.2.3.

The spline filter is faster than the two-dimension FFT. It requires $O(N)$ additions and $O(N/k)$ multiplications while the FFT requires $O(N \log N)$ additions and multiplications, where N is the number of pixels in the image and k is the knot spacing. The number of real additions and multiplications that are required for the base-2 FFT are [Brigham]:

$$\text{Real Multiplications: } (2\gamma - 4)N - 4$$

$$\text{Real Additions: } (3\gamma - 2)N + 2$$

where:

$$N = \text{number of pixels in image}$$

$$\gamma = \log_2 \text{ of } N \text{ (where } N \text{ is a power of 2)}$$

The number of real additions and multiplications that are required by the spline filter were determined to be:

$$\text{Real Multiplications: } 132 \cdot (N^{1/2} - 1)(86 + 46/k) + (N^{1/2} - 1)^2(32/k)$$

$$\text{Real Additions: } 144 + (N^{1/2} - 1)^2(20/k + 16/k^2 + 4) \cdot (N^{1/2} - 1)(86 + 12k + 46/k)$$

where:

$$k = \text{knot spacing; } 1 < k \leq N/4.$$

Note that for a two-dimensional FFT, twice as many multiplications and additions are required (for forward and reverse transforms) as well as at most N multiplications for the filtering operation.

B-splines are used to interpolate the spline image. The interpolated image is composed of piecewise continuous polynomials that are essentially linear combinations of the B-spline basis functions. Knot points define the guiding polygon, a convex hull under which the interpolated function is formed. The variant diminishing property of the spline functions assures that the interpolated function will always lie beneath the convex hull that is defined by the guiding polygon. B-spline basis functions have the property of local support, which permits the positioning of the knots to have local control, see Fig. 2.2.1. That is, if the position of a knot were perturbed the shape of the interpolated function would change only in the vicinity of that knot. The spline filter is spatially variant, unlike the FFT which is spatially invariant, due to the local support property of the spline basis functions. The general equation for a B-spline curve is [Wu et. al.]:

$$P(u) = [x(u), y(u)] = \sum_{i=0}^M B_{i,M}(u) V_i \equiv \sum_{i=0}^M B_{i,M}(u) [V_i^x, V_i^y]^T$$

Where $B_{i,M}(u)$ is the i -th basis function, a compound polynomial of order M . This polynomial is continuous up to and including the $(M-2)$ -th derivative. The degree of the polynomial is $M-1$. The following equation is a simplification of the above for cubic, $M=4$, B-splines; the type used in this work.

$$P_i(S) = [S^3 \ S^2 \ S \ 1][C][V_{i-1} \ V_i \ V_{i+1} \ V_{i+2}]^T$$

$$[C] = 1/6 \begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 0 & 3 & 0 \\ 1 & 4 & 1 & 0 \end{bmatrix}$$

where:

- C = a matrix of coefficients of the periodic uniform B-spline basis functions.
- $i = \{0, 1, \dots, m\}$ where $m+1$ is the number of spans associated with the guiding polygon

2.2 - Preprocessing: Spline Filtering and Histogram Equalization

which has $m+1$ sides and $m+1$ vertices (V_0, \dots, V_m)

$$S = (u - u_i) / (u_{i+1} - u_i); \quad S \in [0,1].$$

Generally, B-spline functions are used to interpolate continuous surfaces, as in computer graphics. These shapes are usually closed curves. However, the spline filter requires splining of an open curve. For an open B-spline curve two end vertices, V_{-1} and V_{M+1} , are extrapolated. See Fig. 2.2.1 for an illustration of splining on an open curve.

$$\begin{aligned} V_{-1} &= 2V_0 - V_1 \\ V_{M+1} &= 2V_M - V_{M-1} \end{aligned}$$

The equations of the new ends are:

$$\begin{aligned} P_0(0) &= (1/6)(V_{-1} + 4V_0 + V_1) \\ P_{M-1}(1) &= (1/6)(V_{M-1} + 4V_M + V_{M+1}). \end{aligned}$$

While testing the spline filter I noticed that the edges of the lungs had a splotchy appearance and the lung area lacked detail, Fig. 2.2.2. This splotchy appearance was attributed to both the effect of the discontinuity at the lung border and to the effect of the significantly darker non-lung area on the interpolated image. Presumably this appearance is due to the interpolated image under-approximating the original at the border. The amount of under-approximation is dependent on the closeness of a knot to the edge of the lung. The pixels values in the non-lung area are set to the mean pixel value of the lung regions prior to splining; this reduces the noticed effect, Fig. 2.2.2.

A splined image with knot spacing k is generated in two steps. First, every k -th row of the image is splined and then each column is splined. The values used in splining the columns are those values that were interpolated when splining the rows. The endpoints of the rows and columns are obtained by a weighted extrapolation of the neighboring knot values; for example, for the left side: $V_{[-1,y]} = V_{[0,y]}/2 + V_{[1,y]}/3 + V_{[2,y]}/6$ and for the bottom left corner: $V_{[-1,-1]} = (V_{[0,1]} + V_{[1,1]} + V_{[1,0]}) / 3$. The region outside the lung parenchyma, the non-lung region, is set to the mean value of the lung region prior to splining.

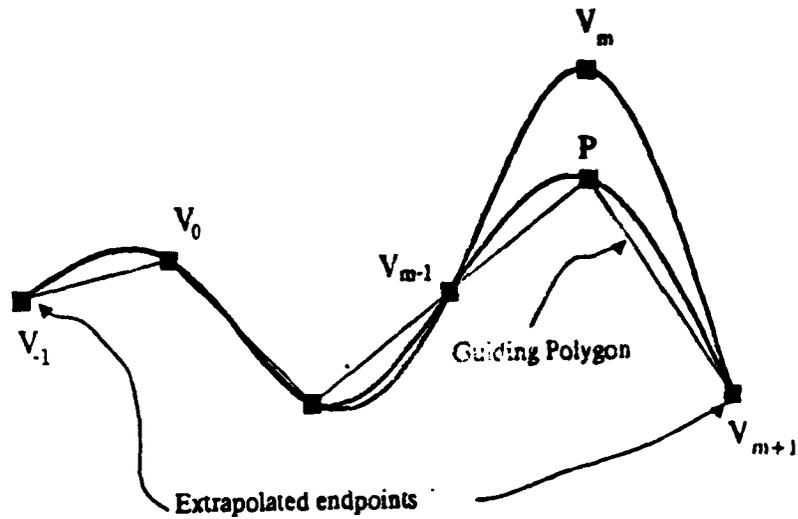


Figure 2.2.1 - Splining on an open curve. The endpoints of the guiding polygon for the open curve are extrapolated from neighboring points. Perturbation of point P on the guiding polygon effects the shape of the curve near P. The interpolated spline curve always lies within the convex hull defined by the guiding polygon. This is a consequence of the variant diminishing property of the interpolation scheme.



Figure 2.2.2 - The effect of setting the non-lung region to the mean value of the lung region prior to spline filtering. The image on the left was spline filtered after the non-lung region was set to the mean pixel value of the lung area. No changes were made to the values in the non-lung area when spline filtering the image on the right.

2.2 - Preprocessing: Spline Filtering and Histogram Equalization

Histogram equalization is a method of expanding the contrast of an image. A cumulative frequency histogram of pixel values, $T(r)$, intensity is determined from the frequency distribution of pixel values, $p_r(r)$, in the following way:

$$T(r) = \sum_{w=0}^r p(w)$$

where, $r \in \{0, \dots, \text{maximum pixel value}\}$.

The histogram equalization, $E(r)$, of pixel value r is given by:

$$E(r) = (r - P_{\min}) / (P_{\max} - P_{\min}) * T(r) / T(P_{\min}) * P^*$$

where, P_{\min} = the minimum pixel value represented in the pixel value histogram,
 P_{\max} = the maximum pixel value represented in the pixel value histogram,
 P^* = the maximum pixel value in the histogram equalized image.

In order to facilitate discussions of operations on the image, the following notation will be used throughout:

$$\Gamma_n[x<:X>, y<:Y>]$$

Γ represents an image array composed of n -bit pixels, X pixels per scanline, and Y scanlines; $\Gamma \in \{A, \dots, Z\}$; the field $\langle * \rangle$ is optional; the domains of the indices are:

$$0 \leq x < X, x = \{0, \dots, (X-1)\}$$

$$0 \leq y < Y, y = \{0, \dots, (Y-1)\}$$

and the range of the image array is:

$$0 \leq \Gamma_n[x, y] < 2^n, \Gamma_n[x, y] = \{0, \dots, (2^n - 1)\}.$$

$$f(G[x, y], p_1, \dots, p_n)$$

f is a function defined in the domain of the image, $G[x, y]$,

p_1, \dots, p_n are the parameters of f .

The spline filtered histogram equalized image, F , is produced from the input image, I , by the spline filter function, f , whose parameter is knot spacing, k .

$$F_g[x:X, y:Y] = f(I_g[x:X, y:Y], k)$$

Knot spacing, the only parameter of the filter, was tuned to provide optimal detection of the known

nodules in the films in the database.

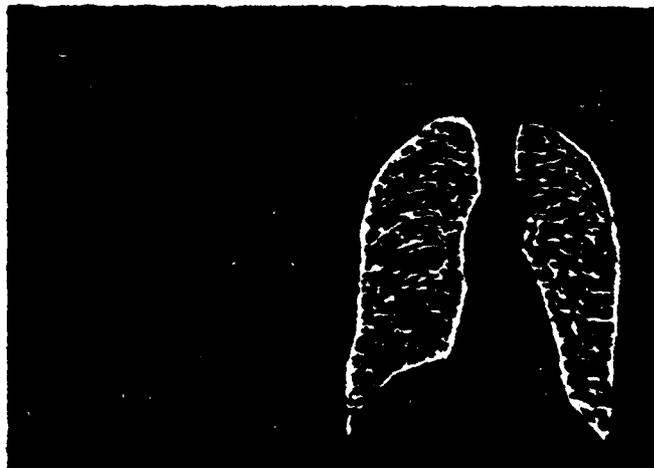


Figure 2.2.3 - Effects of knot spacing and histogram equalization on a spline filtered image. The images on the left were spline filtered at two different knot spacings and were not histogram equalized. The images on the right correspond to those on the left but were histogram equalized. The top images (size: 410 scanlines by 470 pixels per scanline) were spline filtered at a knot spacing of 5 and those at the bottom were filtered at a knot spacing of 120.

2.3 - Candidate Nodule Detection

Candidate nodule detection has three steps: locating the CN center with the CN Expert, image smoothing to accumulate the votes for CN centers, and searching for a specified number of CN centers. The CN detector reports the locations and values of the closed shapes in the image. The value associated with the reported CN center is a function of its edge gradient magnitude and its size. This value is computed by a Hough-like technique [Ballard, 1973]. The essence of the CN Expert is a circle detector which uses embedded knowledge about the appearance of a nodule like shape. The knowledge used by the CN Expert is: that the CN is a closed convex shape that is lighter than its surround. This knowledge is used by the circle detector to determine the location of the center. A simple Hough circle-center locator is used in conjunction with image smoothing by convolution with a Gaussian function to provide a robust CN detector; it is sensitive to a variety of closed shapes, not just circles.

Since the CN Expert is both compute bound and operates on a large (approximately 1 Mbyte) image, it has been designed to minimize the size of its resident set. When many users are on the system, large programs such as this one are swapped. This causes the CN Expert to run slower. The CN Expert, the Gaussian smoother, and the image search operations all operate on horizontal scanlines in a window that moves from the top of the image to the bottom. Only a few scanlines are resident in primary memory at a time. Essentially, the user specifies the number of scanlines that are to be resident in primary memory; these lines are read-in; the next group (a specified number) of scanlines are read-in when a scanline that is above the topmost scanline in the resident window is accessed; access of pixels in scanlines that are below the bottommost scanline in the window (in primary memory) is not possible. This technique has been proven useful in speeding computation time.

The spline-filtered, histogram equalized image, $F[]$, is processed by the CN center locator, $c()$, to produce an image array, $C[]$, that contains the centers of reported CNs.

$$C_g[x:X/resolution, y:Y/resolution] = c(F[x:X, y:Y], radius, resolution)$$

where $c()$ is represented by the following algorithm:

```

for( all image points: y, x in F[]; if edge magnitude > T )
  BEGIN
    cx = x + cos( Edge Angle ) * radius;
    cy = y + sin( Edge Angle ) * radius;
    C[cx, cy] ← C[cx, cy] + 1; /* increment accumulator array at CN center */
  END

```

where,

Edge Angle - the angular orientation of the edge at [x, y] as determined by a Sobel edge operation.

radius - the radius of the sought-after CN; specified by the user.

T - a threshold value.

radius - the radius (in pixels) of the sought-after nodule.

resolution - an integer that specifies the reduction between the dimensions and the input image and the accumulator.

The CN Expert maps the edges of the lighter closed shapes in the image to peaks in $C[x,y]$. That is, the edges of a light convex shape will cast votes via $c()$ in the vicinity of the center of the shape. Convolution with a Gaussian function is used to cluster the votes further about the center of the CN. An integer array of weights is initialized using the following Gaussian function. Two implementational features of the circle detector are that it operates on only a few scanlines at a time and that it is performed using integer weights to minimize floating point overhead. A sparse, non-linear, convolution is performed to restrict the processing to points of probable interest. That is, the pixel value at the center of the convolution template must be greater than a specified threshold if the convolution is to be performed at that pixel.

This array of weights is used to compute $S_{16}[x:X, y:Y]$.

$$S_{16}[x:X, y:Y] = s(C_8[x:X, y:Y], radius, resolution)$$

where $s()$ is given by:

$$S_{16}[x, y] = \sum_{X,Y} \sum C_8[\alpha, \beta] g[x-\alpha, y-\beta] d\alpha d\beta$$

2.3 - Candidate Nodule Detection

radius - the radius of the sought-after nodule in the original image.

resolution - rescales radius; the input image is already rescaled.

$$g[x-x_0: 2*r, y-y_0: 2*r] = \begin{cases} \exp(-\pi[(x-x_0)^2 + (y-y_0)^2]/\text{area}^2), & |x-x_0| \leq r \text{ and } |y-y_0| \leq r. \\ 0, & \text{otherwise.} \end{cases}$$

Following accumulator smoothing the highest, *nvotes*, values and their coordinates in the smoothed accumulator are located and output in a single pass through the smoothed image. As the image is sequentially searched for peaks a circular list is maintained. At the head of this list are the value and coordinates of the largest peak in the image; these was determined by the convolution program during image smoothing. Any image value that is greater than the value of the last item in the list is inserted in the ordered list and the last item is deleted. No insertions are made if the coordinates are within $2r$ and if the new peak value is less than the value of the item already in the list. If the coordinates for a peak whose value, which is about to be inserted in the list, are within $2r$ of an item already in the list and if the new peak value is greater than the one already in the list, that list item is deleted and reinserted in a position appropriate to the new peak value. A $4r \times 4r$ area around each local maximum is set to zero as the peak value and coordinates are inserted in the list. If any local maximum (in the region being set to zero) is encountered that maximum and its coordinates are entered in the list instead. An accumulator list, λ , that contains a specified number, *nPts*, of CNs is the result of searching $S[]$ with the search algorithm, p .

$$\lambda[i] = p(S[x,y], nPts, radius)$$

where $\lambda[i] = \langle a_1, a_2, a_3 \rangle$

$0 \leq a_1 < X$ and $0 \leq a_2 < Y$; where X and Y are the bounds of $S_{16}[x: X, y: Y]$.

$a_3 \in S_{16}[x, y]$.

$i = \{0, \dots, nPts-1\}$, where *nPts* is the number of CNs in λ ,

$a_{\lambda[i]} \geq a_{\lambda[i+1]}, \forall i$.

TR - 120

2.4 - Performance Evaluation

Whether or not nodules are present in the chest film the CN detector will report a specified number of CNs. Ideally, the detector should report any nodules that are present in the film in the highest positions of the list of accumulator peaks. That is, if there are nodules in the film, they should occupy the topmost slots (i.e. have the largest accumulator values) in the ordered list of accumulator votes. Quite often this is not the case. Votes that represent false positives are often interspersed among those that represent actual nodules in the list of CNs. The efficacy of the detector is dependent on the position of the actual nodules in the list of CNs. The cumulative histogram metric (CHM) embodies the following rule: the closer the votes for the actual nodules are to the top of the list and the closer their clustering, the better the performance of the detector. The true positive and false positive rates are used to characterize the performance of the detector. The true positive rate, as it is used in this work, is defined as: the percentage of known nodules that is detected. The notion of false positive, which is somewhat different from the common concept, is: the number of non-nodules that lie between the first accumulator point and the position of the last detected nodule in the list. See Fig. 2.4.1 for an illustration of the calculations of true and false positive rates, and the detection metrics.

A CN is considered a detected nodule if its coordinates are close (a definition of close follows) to those of a known nodule. Forty-four of the digitized films contain at least one nodule (32 contain only one nodule, 12 contain more than one nodule). All of the films in the ANDS database were obtained from Dr. John Wandtke at Strong Memorial Hospital. He specified the locations of the nodules in these films by circling them on an acetate overlay which was placed in register with the radiograph. Later in the computer vision lab, I specified the location of each nodule interactively specified by positioning a cursor over the the nodule in a display of the digital image; an overlay placed in register with its corresponding chest film was used to guide this specification. The locations of the nodules are stored in the header portion of RV (Rochester Vision) images that are 4 or 5 times reductions (per side) of the original image. The criterion that is used by the detection metric to determine if a CN is close to the location of a known nodule is:

if(Dist(accx*scale, accy*scale, knownX, knownY) \leq AllowableError)
 return NODULE DETECTED;

where:

Dist() = Euclidean distance between two points in 2-space.

scale = scaling between accumulator image and image that which contains coordinates of the known nodules.

[accX, accY] and **[knownX, knownY]** = coordinates of CN and known nodules, respectively.

AllowableError = (radius*scale) + ZoomLocError.

radius = radius value that was used by the CN center locator.

ZoomLocError = 2; [*the amount of error (in pixels) allowed when interactively locating the center of the nodule in the reduced image*].

The CHM (Cumulative Histogram Metric) and the true positive rate are used to assess the performance of ANDS. The value of the CHM reflects the placement of nodules in λ . It is defined on [0, 1]. The CHM is the area of the difference between an ideal cumulative frequency histogram, c^* , and the experimentally obtained cumulative frequency histogram, c , (derived from λ) of accumulator votes. The abscissa of this histogram, h , from which the cumulative frequency histogram is derived is the location (actual position) of the detected nodules and the ordinate is $c \in \{0, nNods^{-1}\}$. That is, the presence of a nodule in λ is marked by a delta function with measure (roughly similar to area) $nNods^{-1}$, where $nNods$ is the number of nodules that are known to be in the film.

$$h^*[i] = nNods^{-1}, \text{ if } i < nNods; i = \{0, \dots, nNods-1\}$$

$$= 0, \text{ otherwise.}$$

$$c^*[i] = \sum_{j=1}^i h^*[j]$$

$$h[i] = nNods^{-1}, \text{ if } \lambda[i] \text{ represents a nodule center; } i = \{0, \dots, nNods-1\}$$

$$= 0, \text{ otherwise.}$$

$$c[i] = \sum_{j=1}^i h[j]$$

2.4 - Performance Evaluation

$$c[i] \leq c^*[i], \forall i$$

$$h^*[i], c^*[i], h[i], \& c[i] \in [0, 1].$$

$$CIIM = \frac{\sum_{i=1}^{LastNod} c^*[i] - c[i]}{LastNod}$$

Fig. 2.4.1a illustrates λ , the list of accumulator votes (the statistics above the list of accumulator values were produced by the performance evaluation program). Fig. 2.4.1b illustrates the histogram of accumulator votes and the ideal histogram for an image with 2 nodules. Fig. 2.4.1c illustrates the cumulative histograms derived from the histograms in Fig. 2.4.1b which are used to compute the CHM. A further indication of performance is obtained when the CHM and TP rate are plotted with the CHM as the ordinate and the TP rate as the abscissa, Fig. 2.4.1d. This metric DM, distance metric, is the distance between the TP rate and the CHM and the point of ideal performance, [1, 0]. It is a simple Euclidean distance:

$$DM = \sqrt{(1-CIIM)^2 + TP^2}$$

TR - 120

```

0.5370 ; >>>> DISTANCE Metric for /u/bill/thesis/metric/preAI/191r5k30z2p4. <<<<
0.5370 ; >>>> Cum Misto Metric <<<<
25.0000 ; Number of false positives - # of points which are not
; nodules, that lie between the first accumulator point and
; the last detected nodule.
1.0000 ; Percentage of the 2 known nodules which were detected.
50 ; Number of points in accumulator.
0.5000 ; Percentage of positives in 1-th group of 10 accumulator points.
0.0000 ; Percentage of positives in 2-th group of 10 accumulator points.
0.5000 ; Percentage of positives in 3-th group of 10 accumulator points.
0.0000 ; Percentage of positives in 4-th group of 10 accumulator points.
0.0000 ; Percentage of positives in 5-th group of 10 accumulator points.
0. Acc[141, 142] = 9792
1. Acc[132, 200] = 9504
2. Acc[277, 278] = 9120
3. Acc[120, 222] = 8896
* 4. Acc[353, 117] = 8544
5. Acc[145, 292] = 8064
6. Acc[273, 200] = 7840
7. Acc[283, 250] = 7648
8. Acc[305, 236] = 7648
9. Acc[148, 190] = 7648
10. Acc[291, 285] = 7616
11. Acc[105, 120] = 7328
12. Acc[258, 283] = 7264
13. Acc[155, 288] = 7264
14. Acc[132, 215] = 7232
15. Acc[164, 163] = 7008
16. Acc[369, 221] = 6944
17. Acc[278, 215] = 6880
18. Acc[368, 209] = 6848
19. Acc[129, 164] = 6848
20. Acc[367, 126] = 6816
21. Acc[130, 111] = 6816
22. Acc[147, 165] = 6784
23. Acc[153, 127] = 6688
24. Acc[299, 137] = 6624
25. Acc[188, 332] = 6592
* 26. Acc[ 69, 113] = 6592
27. Acc[284, 320] = 6560
28. Acc[124, 241] = 6560
29. Acc[163, 224] = 6528
30. Acc[183, 352] = 6400
31. Acc[253, 245] = 6400
32. Acc[173, 246] = 6368
33. Acc[ 58, 176] = 6304
34. Acc[321, 254] = 6240
35. Acc[297, 281] = 6240
36. Acc[ 58, 217] = 6208
37. Acc[304, 177] = 6208
38. Acc[256, 216] = 6176
39. Acc[301, 202] = 6176
40. Acc[317, 187] = 6176
41. Acc[258, 281] = 6144
42. Acc[367, 189] = 6144
43. Acc[292, 151] = 6144
44. Acc[186, 306] = 6080
45. Acc[175, 285] = 6080
46. Acc[ 99, 187] = 6080
47. Acc[371, 81] = 6016
48. Acc[ 56, 157] = 5952
49. Acc[106, 355] = 5920

```

Figure 2.4.1a - The list of accumulator peaks that is produced after searching the smoothed accumulator. The nodules that were detected are indicated by an asterisk; 2 nodules are known to exist in the film from which this list was derived.

2.4 - Performance Evaluation

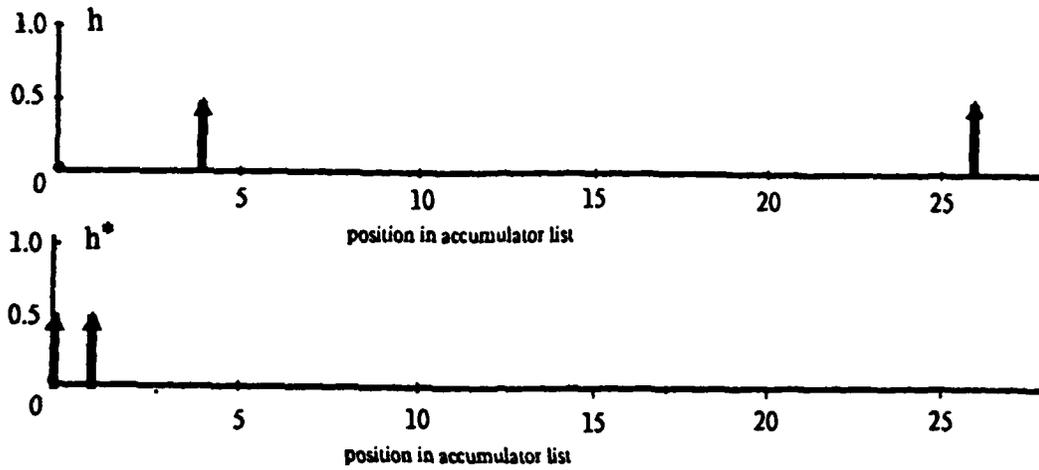


Figure 2.4.1b - The histogram, h , derived from the accumulator list in Fig. 2.4.1a and the ideal histogram, h^* , for a film with 2 nodules.

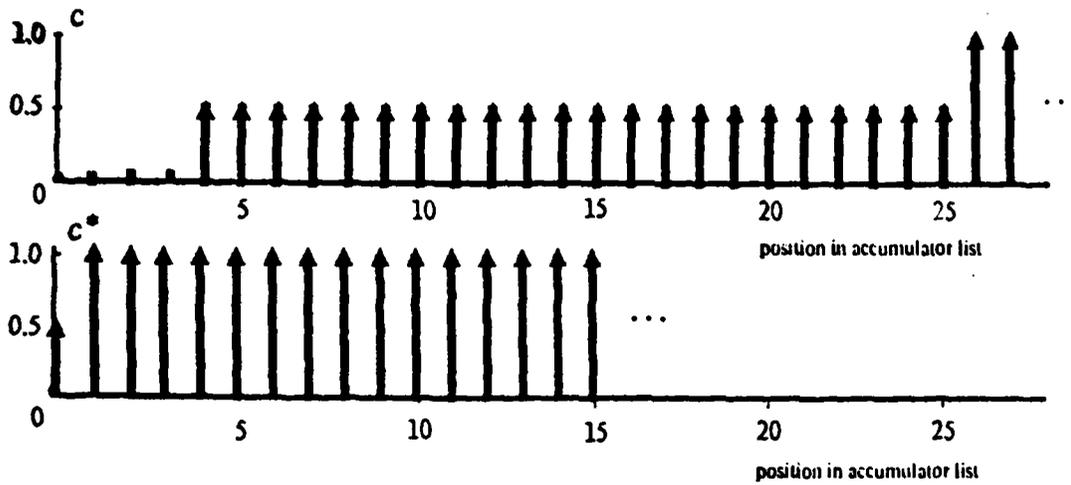


Figure 2.4.1c - The cumulative histograms of the histograms in Fig. 2.4.1b and the CHM which is computed as the area under the difference of these histograms.

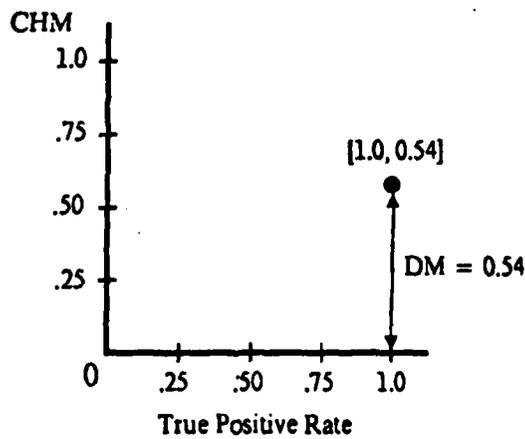


Figure 2.4.1d - A plot of CHM versus true positive rate from which the distance metric, DM, is derived.

TR - 120

2.5 - Incorporation of AI to Reduce the Number of False Positives

Artificial intelligence techniques have been incorporated into the ANIDS with the goal of reducing the false positive rate. Fig. 2.5.1 shows a display of 64x64 windows of the top 16 CNs in λ . Several of the CN images are clearly not nodules. These CNs are false positives. A pattern classifier was taught to recognize the following eleven classes of CNs: Distinct Rib (DR), Small Nodule on Rib (SR), Small Vascularity (SV), Large Vascularity (LV), Small Nodule (SN), Medium Nodule (MN), Large Nodule (LN), Lateral Border (LB), Medial Border (MB), Small Nodule on Border (SB), Nipple (NI), and Undetermined (UD). The incidences of each of these classes are given in Table 2.5.1; these were derived from the classifications of all trained (that is, the classifications were explicitly taught) films.

CLASS	% OF ALL CNS
Rib*	7.8
Small nodule on rib	0.5
Small vascularity*	17.8
Large vascularity*	8.2
Small nodule	1.1
Medium nodule	1.9
Large nodule	0.4
Lateral border*	9.0
Medial border*	21.9
Small nodule on border	0.4
Nipple	0.5
Undetermined*	30.5

Table 2.5.1 - The incidence rates of CN classes. Classes that are considered false positives are indicated with an asterisk. These percentages were derived from all taught CNs; the CNs in all 50 films, each processed at two radii (5 and 10 pixels), which were taught, that is, individually classified by a trained human. Not all CNs were explicitly classified because 64x64 windows centered around the CN could not be made (because the CN is too near the image border), or because the nodule statistics could not be computed (because the CN has a strange appearance). These data come from 2,750 CNs. The CNs that do not fit well into any class are taught as Undetermined. Note: the pattern classifier does not classify nodules as Undetermined; this classification was instituted so that ambiguous CNs would not be used to train the pattern classifier.

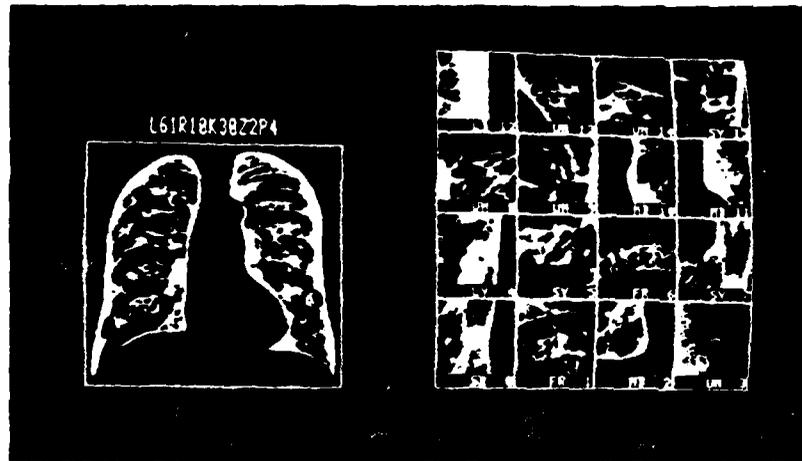


Figure 25.1 - The top 16 CNs from an accumulator list. Several of the displayed CNs are not nodules. Image #0 is a nodule, #1 a rib, #2 a medial border, #s 4 & 5 vascularity, and #12 a lateral border. The classification of each false positive is given in the bottom of the window. These are the classifications that were presented to BMDP7M (a commercial statistical package for linear discriminant analysis) when training the pattern recognizer.

Artificial intelligence techniques have been incorporated in ANDS in the *Nodule Expert*. The *Nodule Expert* is essentially a pattern classifier [Duda, Hart, 1972] with a set of classification rules. These rules determine if a CN is to be omitted from the list of CNs that are presented to the radiologist. This rule causes omission of everything that is not classified as a nipple or some kind of nodule. The pattern classifier uses features which describe the appearance of a CN; the output of two vision experts, the *Rib Expert* and the *Vascularity Expert*; and the position of the CN in the radiograph to classify the CN.

The *Rib Expert* is based on the Hough technique for line detection. Its input is an image that is a windowed region around the center of a CN. This image is histogram equalized and smoothed (high frequency components and noise are removed). The rib expert uses embedded knowledge about the appearance of the rib for guidance as it attempts to reject or accept the image as that of a rib. Salient features of the sought-after object are incorporated into this vision procedure. The following features that characterize *ribbiness* are embedded in the *Rib Expert* algorithm:

- a rib is a light object bounded by two parallel edges;
- by convention of the Sobel edge operator, the angular orientation of the rib edges are separated by 180 degrees;

2.5 - Incorporation of AI to Reduce the Number of False Positives

- the width of the rib is approximately the diameter of the sought-after nodule;
- the rib edges are approximately centered around the center of the CN;
- the parallel rib edges are the strongest (gradient magnitude) of all edges near the CN;
- these edges are also the most extensive edges in the (windowed) image.

The Rib Expert is procedural in that, given the embedded description of the rib, it iterates having the goal of accepting the window image as a rib (segment), rejecting it as a rib, or failing to accept it as a rib. Knowledge about the salient rib features is embedded in both the control structure and the body of the executing statement of the rib expert. An increasing number of image edges are considered as possible rib edges in the control loop. The executing statement tests these edges at two levels. First, a test is performed that attempts to reject the CN as a rib. If that test does not reject the image, the second test attempts to accept the CN as a rib. If neither test is passed then the rib expert iterates further, considering more edges as possible rib edges. The procedural iteration fails to accept an image as a rib only after the top 20% of all edges in the image have been considered as possible rib edges and no rib was yet detected in the image; this is the stopping condition. This stopping condition is based on the notion that the rib edges are the strongest and most extensive in the image.

A Hough transform for a line is computed separately at each iteration for edges that may constitute the top and bottom edges of a rib. Since a rib is essentially horizontal, the top and bottom edges are easy to compute. Each edge, with angular orientation Φ , whose gradient magnitude is greater than the specified threshold is considered in the Hough transform, h , whose output, $H[0:360, \rho:4*\text{radius}]$, represents lines that are described by the parametric equation:

$$\rho = (x - x_0)\cos(\theta) + (y - y_0)\sin(\theta)$$

where,

$[x_0, y_0]$ is the computed center of the CN,

$$\theta = 90^\circ + \Phi$$

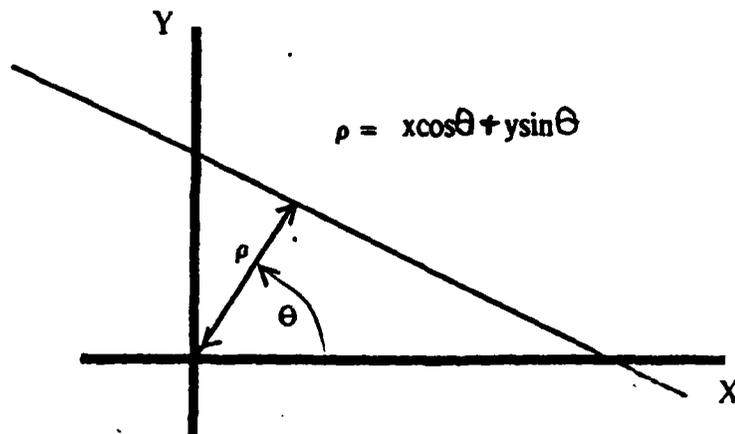


Figure 2.5.2 - The parametric line that is represented by the Hough line transform.

This parametric line is illustrated in Fig. 2.5.2. For each edge, parametric distances, ρ , are computed over a small range of θ ($\pm 5^\circ$), and $H[\theta, \rho]$ is incremented at the appropriate coordinates. Thus, the most predominant line(s) in the image will be represented by the highest valued coordinates in H . After all edges have incremented the parameter space, two histograms, h_1 and h_2 , are derived from the parameter space image; these are frequency histograms of the angular orientation of edges.

$$h_1[\theta] = \sum_{\rho=1}^{4\pi} H[\theta, \rho], \text{ for } 0 < \theta \leq 180$$

$$h_2[\theta] = \sum_{\rho=1}^{4\pi} H[\theta, \rho], \text{ for } 180 < \theta \leq 360$$

These frequency histograms are normalized. These normalized frequency histograms are subjected to peak detection to determine whether or not the image is that of a rib. The image is rejected as a rib if there is more than one peak in either histogram that is greater than 55%. An image is accepted as a rib only if there is one peak in each histogram greater than 55% and if these peaks are within $180 \pm 15^\circ$. A similar procedure is done over ρ . If both tests, θ and ρ , are passed then the image is accepted as that of a rib.

2.5 - Incorporation of AI to Reduce the Number of False Positives

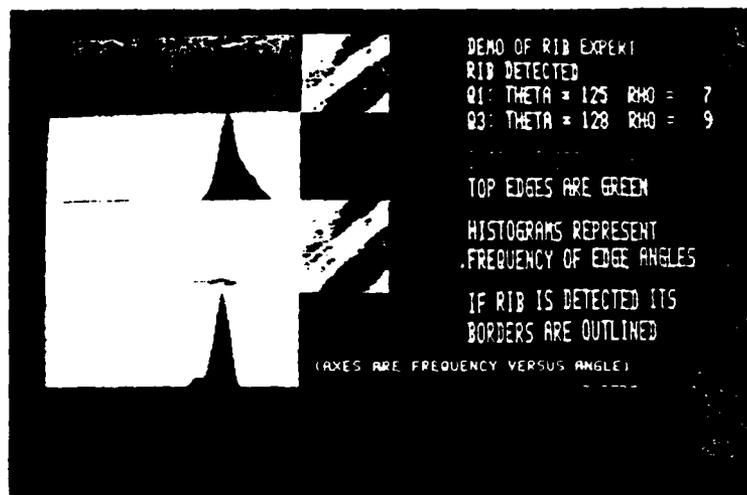


Figure 2.5.3 - The result of running the Rib Expert on a CN that is a rib. The red and green areas on the left are the Hough transform space for the top and bottom edges of the rib, respectively. The abscissae of the histograms below each is angular orientation. The ribs are detected because there are only two such peaks and the peaks are within $180 \pm 5^\circ$ of each other. The red and green lines that are drawn along the rib edges correspond to the top and bottom rib edges that were detected by the Hough transform.

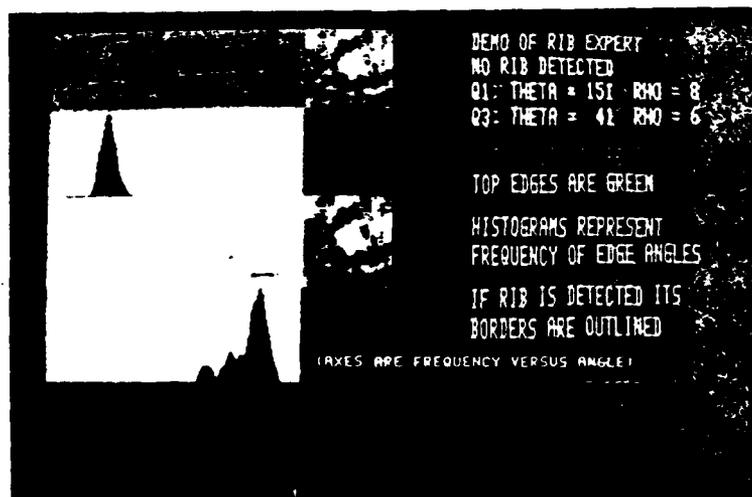


Figure 2.5.4 - The result of running the Rib Expert on a CN that is a nodule. The nodule is rejected as a rib by the detector because the peaks that represent the strongest edges are not $180 \pm 5^\circ$ apart.

In summary, the Rib Expert uses knowledge of the appearance of the rib to detect a rib. The embedded knowledge determines the process through which the data are reduced. Figs. 2.5.3 and 2.5.4 illustrate the results of running the Rib Expert on a rib, which is recognized as such, and on a nodule, which is rejected as a rib. Angular orientation and brightness constrain the rib edge to which an edge element might belong. The rib model, which requires parallel and extensive rib edges, constrains the allowed angular orientation and requires that the preponderant edges in the image fit the model of rib edges. These appearance constraints are used in conjunction with a line detection technique and some procedurally embedded rules to reduce rib detection to peak-finding in 1-dimensional arrays. The effectiveness of the expert depends on the threshold levels used in the peak-finding operations as well as the the sophistication of the embedded knowledge. In preliminary tests, 85% of all rib images (19 were tested) were correctly identified; no false positives were reported.



Figure 2.5.4 - The Vascularity Expert. A back-projected Hough transform is used to measure colinearity clustering of vascular structures. The rectangular regions near the mediastinum in both lungs delimit the areas that would contain CNs that are considered by the Vascularity Expert. The CNs that are considered as candidate vascular structures are marked by red dots. Lines whose orientations correspond to those of hypothesized anatomic structures are drawn between candidate vascularity sites. The intensity value of each of these lines corresponds to the number of vascularity points that lie on the line and the number of lines that pass through the candidate vascularity site.

The Vascularity Expert provides a measure of colinearity of CNs in a region near the mediastinum of both lungs. The assumptions here are that branching vascular structures near the mediastinum

2.5 - Incorporation of AI to Reduce the Number of False Positives

will present circular shadows where the branches are imaged end-on and that the branching structures are somewhat linear. Rectangular regions near the mediastinum enclose the CNs that are considered in this test, see Fig. 2.5.5. A Hough transform for a line is computed for each CN in these regions; each region has a separate transform space. The Hough transform is computed over a range of angles. This range is constrained by the possible angular orientations of vascular structure in each lung. Thus, a line (in transform space) will receive as many votes as there are CNs lying on it. The results of this transform are then back-projected to determine a value of colinearity for each CN. Each CN, as a result of the back transform, is assigned a weight that corresponds to the total number of points on all lines (colinear CNs) that pass through it.

Linear discriminant analysis is incorporated in the pattern recognizer to classify CNs. BMDP7M (a commercial statistics package for doing linear discriminant analysis, which is available on the Medical Center Computing Facility's DEC-10) was chosen to perform the discriminant analysis. There are two aspects of the pattern recognition process: system training and recognition. In the training phase the pattern classifier is presented with feature vectors that typify the classes of CNs. In this phase the pattern classifier develops a multivariate statistical model of the classes and computes a linear function for classifying CN feature vectors. The pattern classifier was first taught using 295, 24-element, vectors (23 feature values and a classification value) from 9 films (it was later taught with 2750 feature vectors from 37 films); BMDP7M computes a set of weights, W_i , and constants, c_i , which are applied to a CN feature vector, x , to determine the discriminant weights, $d_i(x)$, for the i -th class. Table 2.5.2 describes the CN features that were input to BMDP7M.

$$d_i(x) = x^t W_i + c_i$$

The class with the largest computed discriminant value is the class to which the CN belongs. The set of weights, W_i , and constants, c_i , that are provided by BMDP7M are instantiated in the pattern classifier which is part of the Nodule Expert.

TR - 120

FEATURE	DESCRIPTION
Accumulator Value ^{8, 37}	represents the number of votes that were cast for a CN
Relative Medial Distance ^{8, 37}	shortest distance between CN and medial border; normalized by the sum of the medial and lateral distances
Relative Lateral Distance ^{8, 37}	shortest distance between the CN and the lateral border; normalized by the sum of the medial and lateral distances
Relative Central Distance ^{8, 37}	distance between CN and medial border midpoint normalized by the distance between the medial border midpoint and the top or bottom apex, if the nodule is closer to the top or bottom, respectively
Vascularity Weight	a measure of colinearity of CNs that lie near the mediastinum; this value is proportional to the number of nodules that lie on the line(s) that pass through a given nodule
Average ³⁷ /St. Dev. Pixel Value	azimuthal averages are computed for all octants; these statistics derive from the azimuthal averages (average pixel value of an arc in each octant) per octant
Average ^{8, 37} /St. Dev. Radius	statistics on pixel-value boundary points; radial distance in cm.
Hough Radius ³⁷	radius (in pixels) used by ANDS when searching for CNs
Average/St. Dev. dR at Gradient Boundary	distance between the gradient magnitude peak (in cm.) and 50% of that peak value in a histogram of azimuthally averaged gradient magnitude, per octant; relates to edge contrast
Average/St. Dev. ³⁷ dG at Gradient Boundary	change in gradient magnitude between peak gradient magnitude and 50% of that peak; relates to edge contrast
Average/St. Dev. ^{8, 37} Edge Strength	ratio between dG and dR (above) over all octants; relates to edge contrast
Average ³⁷ /St. Dev. ^{8, 37} Edge Visibility	ratio between the maximum gradient magnitude for each octant and the maximum gradient magnitude over all octants; describes the uniformity of the edge gradient around the CN
Avg. ³⁷ /St. Dev. Change in Int./Ext. Brightness	change between interior and exterior brightness at pixel value boundary, over all octants
Avg./St. Dev. ³⁷ Ratio between Ext./Int. brightness	ratio between average interior and exterior pixel value across the pixel value boundary, over all octants
Rib Expert value ³⁷	value returned by the rib expert; essentially a Boolean value

Table 2.5.2 - Descriptions of the 23 CN features that were input to BMDP7M in the training phase. The super-scripts, 8 and 37, indicate which features were used when training the pattern classifier 8- and 37- trainings, respectively.

2.5 - Incorporation of AI to Reduce the Number of False Positives

The nodule appearance features characterize essentially two aspects of the CN: brightness and margination. The CN appearance features are computed using knowledge about the locations of two types of CN boundary: the pixel value and the gradient magnitude boundaries. The pixel value boundary is determined as a side effect of having computed the nodule center. The CN Expert gives a rough estimate of the center of the CN. A center finder refines the estimate of the center that is located by the CN Expert.

The CN center finder locates 8 points on the CN border. These points are 45 degrees apart, with respect to the CN center. If one of these points cannot be computed then the features of the CN are not determined. Each pixel value boundary point represents the region between the inside and the outside of the CN. Each point is determined separately. The boundary point is essentially the point of inflection of the change in pixel value of adjacent pixels on a radial arm. This point is determined as the peak in a cumulative histogram of the differences between adjacent pixels (on the radial arm) which are indexed by radial distance, see Fig. 2.5.7. The radial arms are the radii which divide the CN into equal-sized octants; one of the arms has an angular orientation of 0° with the horizontal, see Fig. 2.5.6. The average radius, changes between the inside and outside brightness of the nodules, and average pixel value are features that are determined using this boundary. These essentially describe the light/dark properties of the CN and the distinction of the CN from the surround.

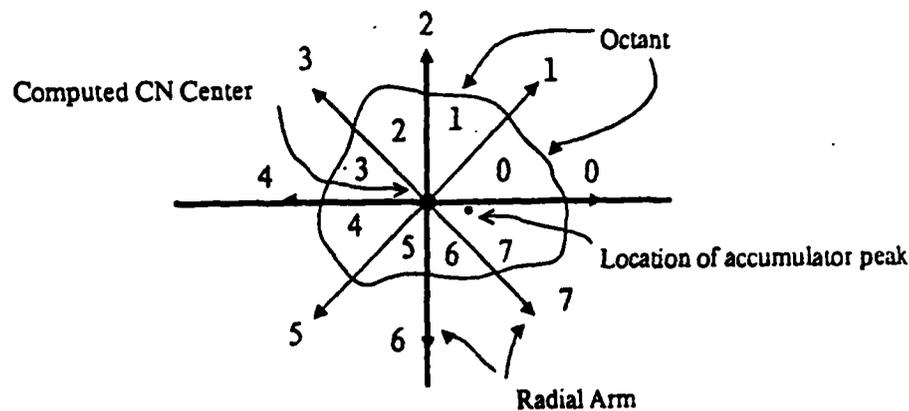


Figure 2.5.6 - The partitioning of the candidate nodule. The CN is divided into equal sized octants by the radial arms. The pixel value boundary point is defined on each radial arm. It delimits the inside from the outside of the nodule based on brightness considerations. The gradient magnitude boundary is defined as a radial distance of an arc in each octant. This boundary delimits the inside from the outside of the CN based on sharpness considerations (per octant)

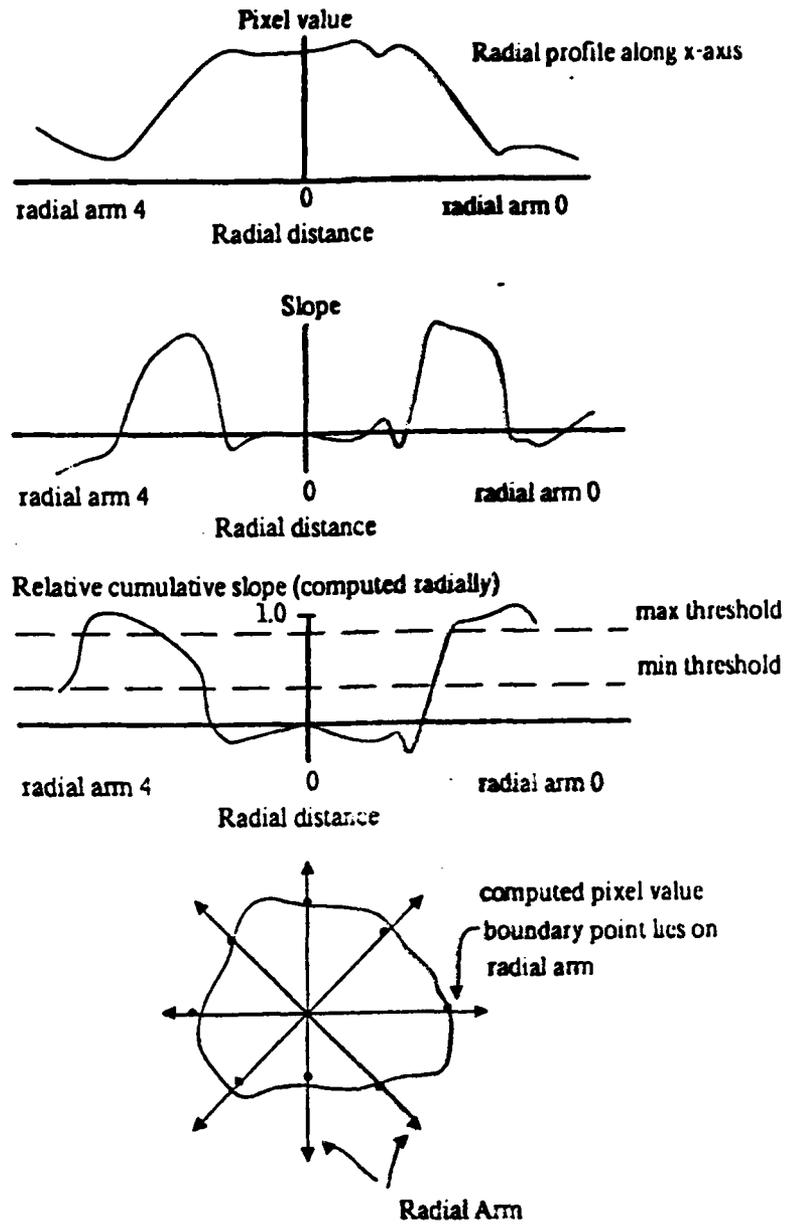


Figure 2.5.7 - Computation of the pixel value boundary. The pixel value boundary is a set of points on the radial arms. Each boundary point is computed with a peak finding procedure, which is illustrated above. First, radial slope is determined along a radial arm. This slope is defined as: $slope = \frac{|x-dx, y-dy| - |x+dx, y+dy|}{|x-dx, y-dy| + |x+dx, y+dy|}$. It is derived in a smoothed image. Next, the relative cumulative radial slope (the normalized cumulative area under the plot of radial slope) is computed. The location of the pixel value boundary is derived from this plot. The boundary point is defined as the first peak greater than a minimum threshold, or the first radial point that is just greater than a maximum threshold.

2.5 - Incorporation of AI to Reduce the Number of False Positives

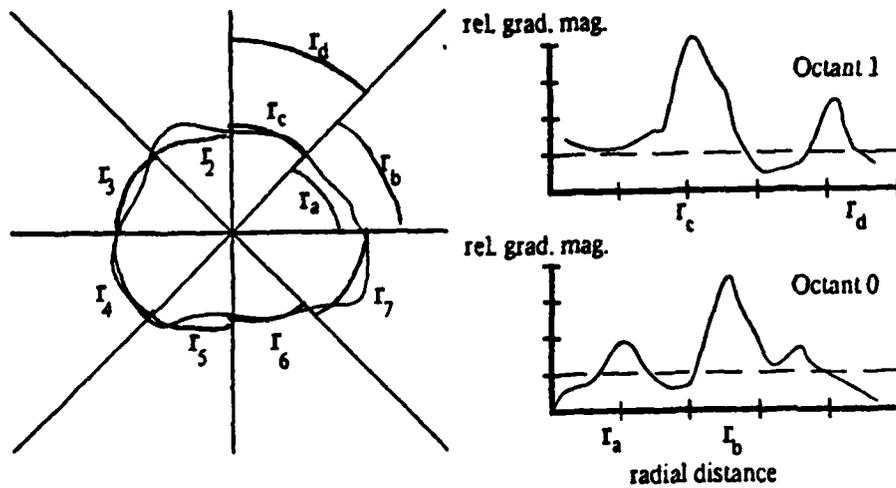


Figure 2.5.8 - Computation of the gradient magnitude boundary. The gradient magnitude boundary is composed of arcs, one per octant, that are derived from peaks in a histogram of azimuthal averages of gradient magnitude. These arcs are illustrated at the top left. Two octants, 0 and 1, contain two arcs. The radial locations of these arcs correspond to the locations of the (at most two) largest relative gradient magnitude peaks for the respective octants. These peaks are pictured in the upper right. All the other octants have only one significant peak in their gradient magnitude histograms (which are not illustrated), since they contain only one arc. A recursive procedure is used to determine the most consistent gradient magnitude boundary. A consistent boundary is one in which the total radial distance between adjacent candidate boundary arcs is minimized. The candidate boundary is represented by the graph on the bottom left. Nodes represent candidate boundary arcs. Edges represent the radial distance between adjacent boundary arcs. The recursive procedure prunes arcs that do not lead to a minimal cost path; this produces a consistent gradient magnitude boundary.

The gradient magnitude boundary is determined in a similar fashion to the pixel value boundary. However, the gradient magnitude boundary (for each octant) is determined from a histogram that is obtained from the azimuthally averaged gradient magnitude in each octant. This boundary is determined as the minimal cost path through the (at most two) peaks in the azimuthally averaged gradient magnitude histograms for each octant, see Fig. 2.5.8. That is, the minimal cost path results in a boundary in which the radial boundary distance per octant is consistent with those of its two adjacent neighbors. The gradient magnitude boundary lies on or beyond the pixel value boundary. For sharp edges it lies on the pixel value boundary and for fuzzy edges it lies radially beyond it. Edge strength, edge visibility, and change in edge gradient are features that are determined using the gradient value boundary. These features describe the definition of the CN margin or its separateness from the surround.

In addition to the nodule appearance features, several relative distance features were devised for use in the pattern recognizer. They complement the knowledge about local features by adding global knowledge about the relative position of the nodule in the image. These relative distances are: central distance (from the middle of the medial border, see Fig. 2.5.11, to the nodule); medial distance (from the medial border to the nodule); and lateral distance (from the lateral border to the nodule). The central distance is normalized by the distance between the medial center and the top of the lung, or the distance between the medial center and the bottom of the lung, depending on whether the nodule is in the upper or lower portion of the lung, respectively (see Fig. 2.5.11).

Specific locations in the lungs must be known so that the relative distance measures can be determined. These locations are the medial and lateral borders, the top and bottom apices, and the medial midpoint of each lung, Fig. 2.5.10. First, the lungs must be located in the image. Locating the lungs is aided by the facts that the non-lung area of the image was masked-out when copying, and that the lung image has the correct orientation. The non-lung region is known to contain pixel values that are lower than any in the lung region. A line midway between the medial borders of the two lungs is located using a projection of the image onto the horizontal axis. This line is used when

2.5 - Incorporation of AI to Reduce the Number of False Positives

computing the lung borders and relative distance features to distinguish the right from the left lung. Successive horizontal lines that lie in the lung parenchyma of each lung are considered when computing the lung borders, Fig. 2.5.9. The endpoints of the longest horizontal line in each lung are taken as the lung borders. Two lung borders arrays, lateral and medial, are computed by the lung border detector for each lung. For example, to determine the lateral border at a given vertical height, the lateral border array is indexed to obtain the horizontal coordinate of that border. The lung border arrays are strictly positive for all vertical coordinates that are in the lung parenchyma.

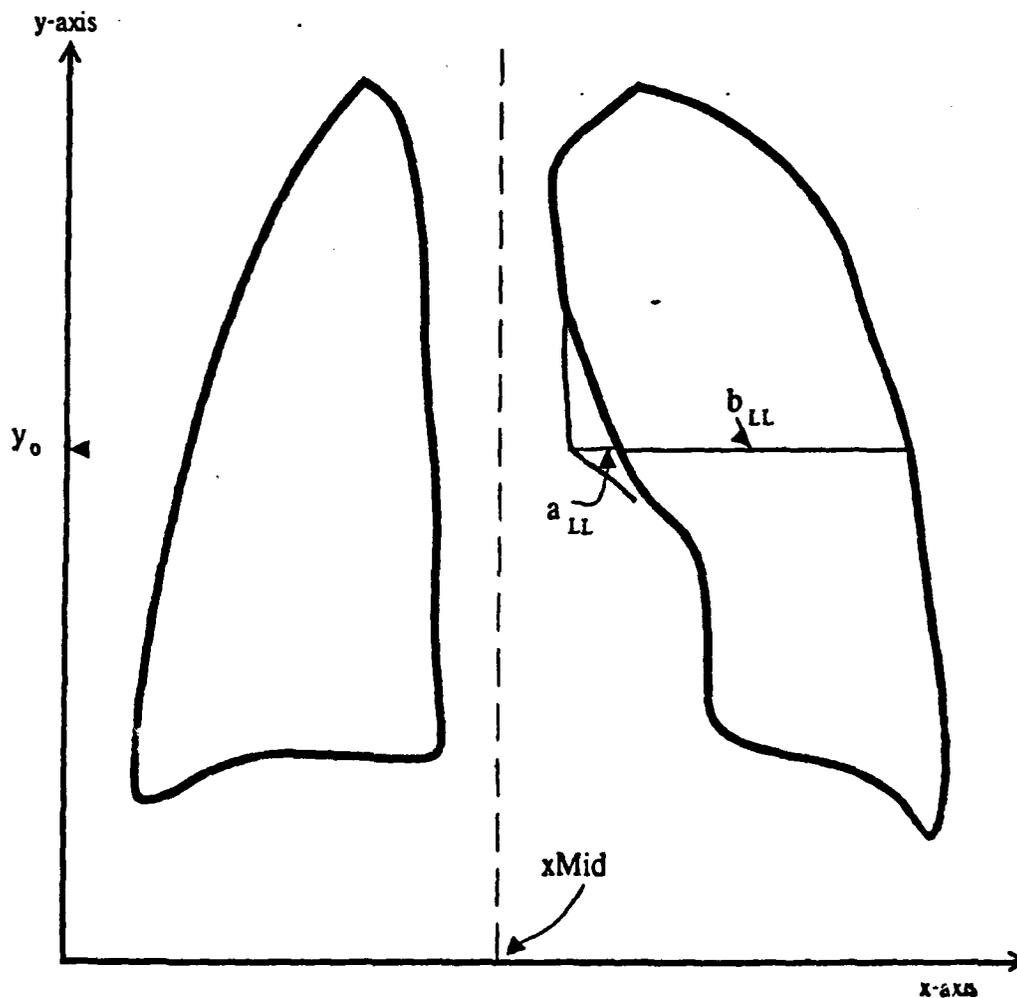


Figure 2.5.9 - Computation of lung borders. The location of each lung in the image is computed and the line, x_{Mid} , which is midway between them is derived. Two arrays (one for the lateral border and the other for the medial border) are computed for each lung. The index of each array is vertical height and the value contained in the array is the horizontal coordinate of that lung border; the value is -1 if the horizontal height is not within the lung. The endpoints of the longest line segment in each lung are taken as the respective lung borders, when computing the lung borders. Two segments are illustrated in the left lung, above. These are likely to arise at the lung border because of anomalous structures. Here, a_{LL} is longer than b_{LL} so its endpoints are determined to be the horizontal boundaries of the right lung at y_0 by the lung border locator.

2.5 - Incorporation of AI to Reduce the Number of False Positives

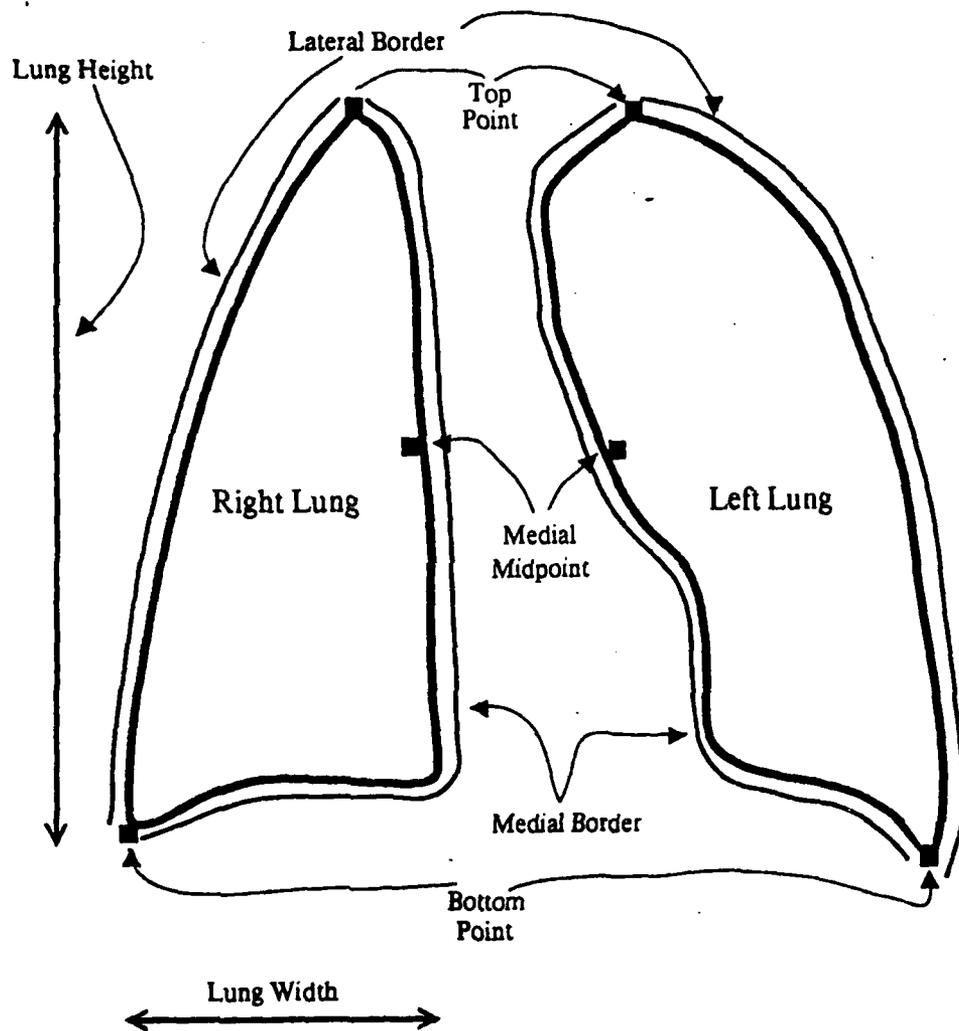


Figure 2.5.10 - Computed lung locations. The lateral and medial borders of each lung are computed by the lung border locator; from these, the heights and widths, and tops and bottoms of each lung are determined by the border locator as well as the medial midpoints of each lung.

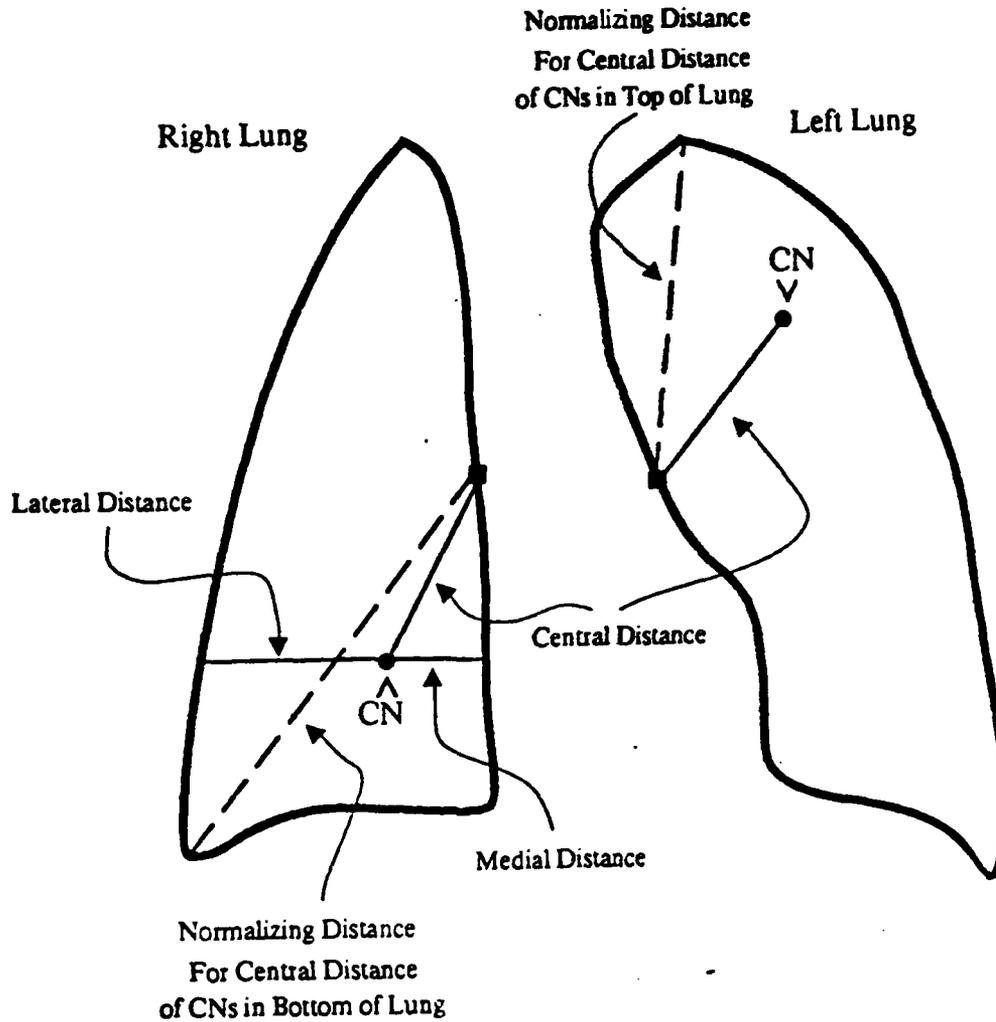


Figure 2.5.11 - Relative lung distance features that are used by the pattern classifier. The relative lateral and medial distances are those distances from the respective lung borders to the CN that are normalized by the width of the lung at the height of the CN. The relative central distance is that distance between the nodule and the medial border midpoint normalized by the distance between the medial midpoint and the top/bottom of the lung, depending on whether the nodule is in the top or bottom half of the lung, respectively.

3 - Experimental

The final ANDS consists of the following steps: spline filter, smoothing of the spline filtered image, detection of CN centers, smoothing, and search for the most prominent CN centers. Four processing configurations were tested on 5 films to determine which configuration would provide optimal detection. The parameters (knot, image/accumulator rescaling, image resolution, and radius) were tested over 5 films to determine which provides optimal detection. This processing configuration and its parameters constitute the basic ANDS.

Artificial Intelligence in the form of a Nodule Expert that uses a pattern classifier and two procedurally driven nodule experts (which detect two classes of false positives) are incorporated in the basic ANDS in order to reduce the number of false positives. Forty-three films were processed by ANDS with and without AI. The results of the two runs are compared.

The overall goal of the experimental work is to determine a nodule detection method that best detects nodules, to tune that process, and then to reduce the number of false positives that are reported by it. The experimental work has four parts: optimization of the linear transfer of optical densities to pixel values during photographic copying and digitization; choosing a nodule detection process; tuning the parameters of the nodule detection process; evaluation of the ability of AI techniques to reduce the number of false positives. A method for photographically reducing and then digitizing the chest radiograph image that is both linear and repeatable was first devised. The careful definition of the photographic reduction and digitization methods permits additional chest films to be added to the current ANDS database without the introduction of batch-to-batch variation. The linear transfer of optical densities to pixel values assures that nodular abnormalities of various densities will be represented without degradation in the digital image.

TR - 120

3.1 - Optimal Reproduction of ANDS Database

Fig. 3.1.1 illustrates the generalized photographic/digitization process. Three transfer functions: camera flare, film characteristic, and digitizer characteristic determine how optical densities from the chest radiograph are transformed into pixel values. Ideally, the system transfer function should represent a linear mapping between optical density and pixel value. The shape of the flare curve is determined by the amount of light that is reflected within the camera system; the linearity of this transfer decreases with increasing amounts of internally reflected light. The shape of the film characteristic is determined by the film type (emulsion) and its development. The placement of the input densities on the linear portion of the film characteristic is determined by the exposure. The shape of the digitizer characteristic is determined by the adjustment of the gain of the A/D converter in the film digitizer.

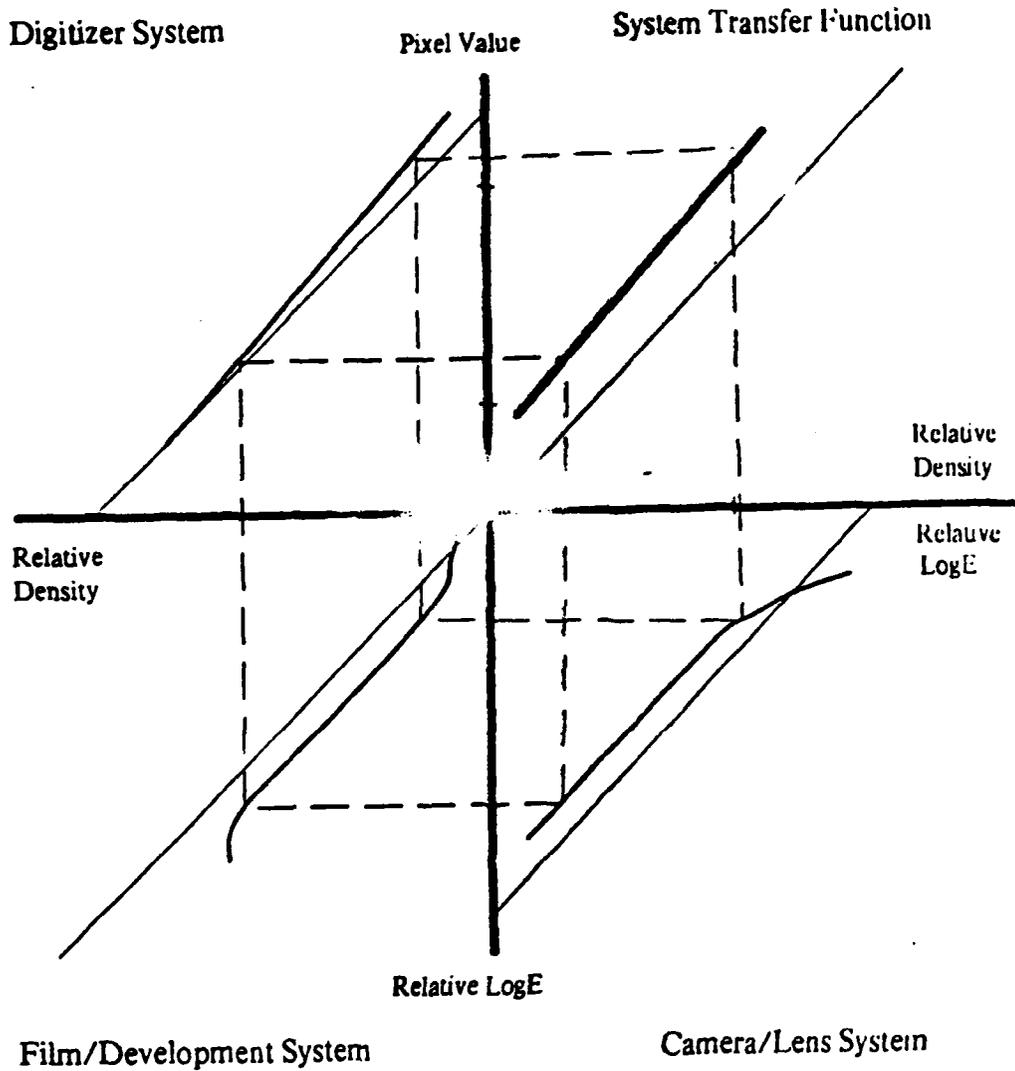


Figure 3.1.1 - The optical density to pixel value transfer function in quadrant I is the result of three cascaded systems: camera/lens, film/development, and digitizer calibration. The dotted lines represent the ideal system transfer functions. A linear system-transfer function is desirable. An optimal system tone reproduction characteristic was experimentally determined. The digitizer and film/development characteristics were preset at optimal levels and the optimal exposure and flare condition were experimentally determined.

3.1 - Optimal Reproduction of ANDS Database

A film, Kodak Commercial, that provides a useful linear range of approximately 3.0 logI and a gamma near 1.0 was chosen for photographic reduction. The development process was fixed as HC-110-D, at $68 \pm \frac{1}{2}^{\circ}\text{F}$ for 5' using R.I.T. tray rock agitation. The digitizer was calibrated with a 5-inch Kodak #2 step wedge to provide the greatest possible linear range with optimum discrimination at high (≈ 2.70) densities.

Given that the Optronics was calibrated to produce a near linear transfer of optical densities to pixel values and that the film development was fixed to produce a gamma near 1.0, the conditions that were varied were exposure and flare condition. Three exposures and three flare conditions were evaluated. These nine exposure/flare combinations were evaluated using three radiographs whose density ranges (in the lung area) typify the population of radiographs that was digitized.

In-camera sensitometry was used to determine the approximate exposure to be used when copying the chest films. The camera system was set up identically when determining this exposure and when copying the 50 radiographs. The approximate exposure was experimentally determined by photographing a 10" Kodak #2 step wedge centered on the light table with the luminous area surrounding the step wedge masked with exposed x-ray film (density $\gg 5$). An exposure that provided a near 1:1 mapping of step wedge densities to developed film densities was chosen. As each chest radiograph was copied the 10" step wedge and a tri-bar target were included alongside the radiograph so that reproduction might be quantitatively assessed.

A discrepancy was noted, however, between the reproduction characteristics derived from the masked step wedge alone and those derived from the same step wedge imaged alongside a chest radiograph. The system transfer characteristics that were derived from film samples which were given identical development and digitized consecutively (no adjustment was made in scanner calibration between runs) are plotted in Fig. 3.1.2. The discrepancy between the characteristics was attributed to camera/lens flare and/or ambient light reflecting from the surface of the radiograph when copying. A tent of black velvet was constructed around the camera system to eliminate ambient light (both room light and light from the light table that was reflected from the ceiling).

Construction of this tent resulted in only a slight decrease in the noted discrepancy. Thus camera/lens flare was determined to be the primary cause.

The source of the camera/lens flare was presumed to be light passing through the non lung area of the chest radiograph, which is noticeably lighter than the imaged lung parenchyma. Since flare was shown to diminish the reproduction of the higher densities, the non lung region was masked with exposed x-ray film. Thus, masking would diminish the amount of light which would reflect within the camera system and lead to degradation of tone rendition of higher densities. Whether masking the non-lung region produced a significant reduction in camera flare is the subject of the following statistical analysis.

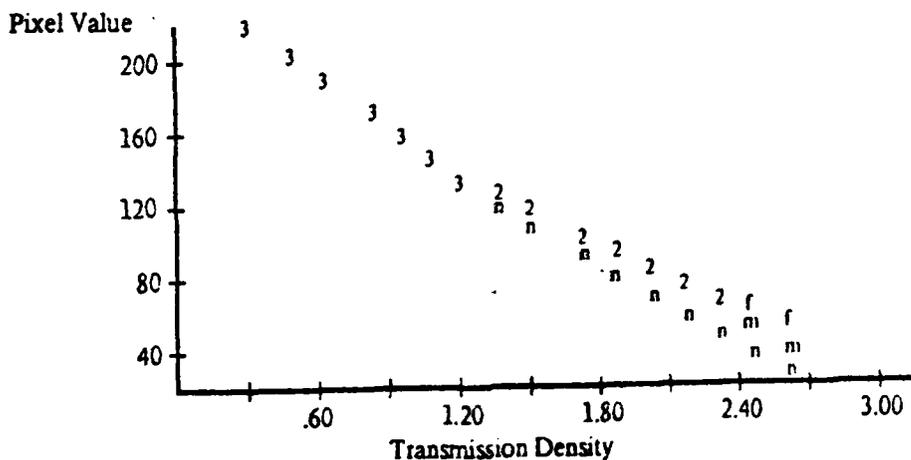


Figure 3.1.2 - System transfer curves of the step wedge for the same step wedge exposed at three flare conditions: n = no flare, f = flare, and m = masked. All samples were developed and digitized under identical conditions. The differences in the toe portion of the curves is presumably due to camera/lens flare.

Three chest radiographs whose density ranges (in the lung parenchyma) were chosen to represent the population of radiographs that was to be digitized. Each of these three radiographs was copied at two flare conditions (masked and not masked) and three exposures (10sec. at f/16, f/22, and f/32). The third flare condition that was evaluated represented the ideal case, no flare.

3.1 - Optimal Reproduction of ANDS Database

only the step wedge was copied. Images that were produced at the three flare conditions are illustrated in Fig. 3.1.3. Weighted averages (weights correspond to the relative percent of the population represented by each film) for each of the 21 steps (in a 5x5 pixel area) were derived from digitized images of the three films at each flare/exposure condition. A second-order regression was computed from these 21 averaged steps for each exposure/flare combination, regressing on pixel value as a function of transmission density.

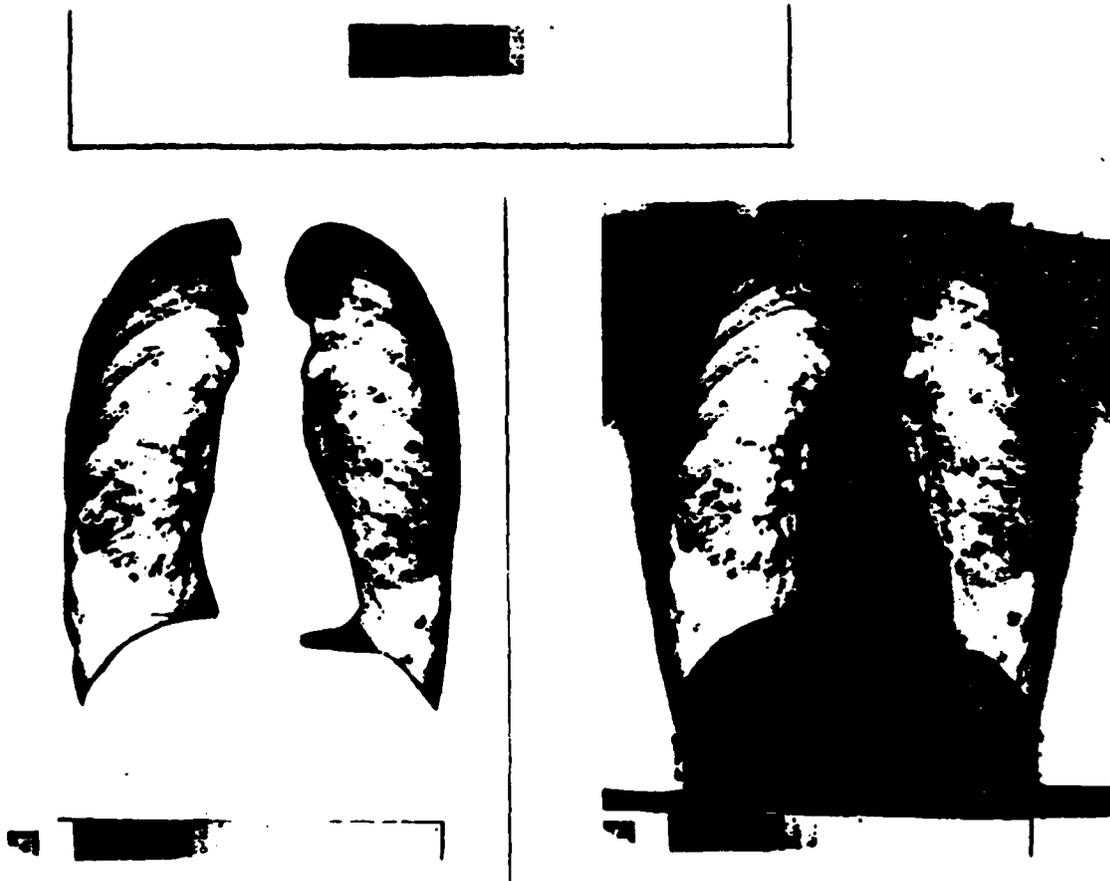


Figure 3.1.3 - Three flare conditions (no flare, top; masked, left; and flare, right) were evaluated to determine which (masked or not masked) would result in the most linear transfer from optical density on the chest radiograph to pixel value in its digital representation. Measurements of the step wedge densities from each of the above images are plotted in Fig. 3.1.2. The objective of this part of the experiment is to determine if masking the non-lung area would provide better density to pixel value transfer because of the presumed reduction in the amount of camera/lens flare.

The goal of the regression was to determine which flare condition, masked or unmasked, provided the most linear tone reproduction. Here, linear reproduction is defined as a lack of statistical significance in the second-order term of the following regression equation:

$$P = a_0 + a_1D + b_1D^2$$

where: P = pixel value,
 D = transmissive optical density,
 b₀...b₂ = coefficients of the regression on D.

The regression analysis was performed using Minitab [Ryan, 1976]. Table 3.1.1 summarizes the regression analysis. This regression was performed over optical densities in the useful range of 0.20 to 2.72; the densities in the lung (11) of the 50 films were within this range. The only exposure/flare condition that resulted in a statistically significant second-order regression term is masked at f/22. No trend was observed in the plot of residuals from the regression equation; this suggests that the proposed model sufficiently represents the data. Masks were cut for each of the 50 films; each was copied at 10 seconds at f/22; given identical development; and digitized on the calibrated scanner. The calibration of the digitizer was periodically checked by digitizing the #2 step wedge. Calibration was maintained throughout digitization, no recalibration was required.

f-stop	Student's t-statistic		
	Flare Condition		
	No Flare	Masked	Flare
16	-1.61	-3.47	4.27
22	0.58	1.28*	6.83
32	0.36	5.62	9.34
d.f.	18	13	13

Table 3.1.1 - Student's t-statistics on the second-order term of regression analyses of pixel value as a function of optical density, quadrant I of Fig. 3.1.1 for three exposures and three flare conditions. As a result of this analysis masked and f/22 were used when copying the 50 films in the ANDS database. This exposure flare condition was the only practical condition in which the second-order term of the regression is statistically insignificant.

3.2 - Choosing an Image Processing Configuration

Four image processing configurations were tested with their parameters set at fixed levels to determine which resulted in optimum detection of the pulmonary nodules present in 5 films. The tested configurations are:

- #1 Spline filter with histogram equalization
Spline smoothing
Candidate nodule detection
Spline smoothing
Vote accumulation
- #2 Spline filter with histogram equalization
Spline smoothing
Candidate nodule detection
Sparse convolution smoothing
Vote accumulation
- #3 Spline filter with histogram equalization
Candidate nodule detection
Spline smoothing
Vote accumulation
- #4 Spline filter with histogram equalization
Candidate nodule detection
Sparse convolution smoothing
Vote accumulation

The differences between these methods are spline smoothing following spline filtering and spline or convolution smoothing following candidate nodule detection. All five films (6, 18, 32, 26, 41) were processed at the same parameters (resolution: no rescaling between the filtered image and the accumulator array of candidate nodule centers; radius = 12 pixels; knot spacing = 60, and no rescaling of the original image (size $\approx 1000 \times 1000$ pixels). A weighted sum of three measures: true positive percentage, false positive fraction (the ratio of the number of non-nodules between the first accumulator point and the last detected nodule and the number of points in the accumulator list), and the DCHM is used in a two-way ANOVA in which the processing configurations are treatments and the films are blocks. Where the DCHM is similar to the CHM but is coarser:

$$\text{DCHM} = \sum_{q=0}^Q (1 - 0.1q)h(q)$$

where:

Q = number of quantiles. Here, there are 10 quantiles, each containing 10 accumulator points.

h(Q) = the percentage of all detected nodules per quantile, Q.

A summary of this ANOVA is given in table 3.2.1. The data and ANOVA calculations are presented in Appendix 9.2.

Source	Sum of Squares	df	Mean Square	F-ratio
Treatments	0.0415	3	1.383×10^{-2}	2.01
Blocks	0.9690	4	0.2423	35.19*
Residuals	0.0826	12	6.885×10^{-3}	
Total	1.0931	19		

Table 3.2.1 - ANOVA results for processing configurations as treatments and radiographs as blocks. Data and calculation of ANOVA are given in Appendix 9.2. The effects of film are significant at $\alpha=0.05$ ($P < 0.001$). The variation due to blocks, which is a result of the non-homogeneity of the films, may have obscured any differences due to treatments.

The contribution of the treatments (processing configurations) to the total variance is not statistically significant at the $\alpha=0.05$ level, while the contribution of the blocks (films) is statistically significant at $\alpha=0.05$ ($P \approx 0.001$). The amount of variation among blocks obscures any differences due to treatments because of the non-homogeneity of the films. Here, non-homogeneity implies structures represented in the films, their relative sizes, shapes, and intensities, which vary among films. One may not draw a conclusion founded on these statistical analysis that the differences among the processing configurations are significant.

Processing configuration #4 was chosen for incorporation in ANDS. The table of data from which the ANOVA was computed (see Appendix 9.2) indicates that methods #1 and #3 have the lowest levels of performance (with means of 0.71 and 0.68, respectively) over all 5 films. Configurations #2 and #4 have the highest levels of performance (means of 0.79 and 0.76, respectively). The choice is between methods #2 and #4. Although method #4 does not have the highest performance, it was chosen as the basis of ANDS because it was close to #2 (within 4%) and because it requires one less step (spline smoothing following spline filtering). Also configuration #4 is faster, requiring one less step than configuration #2. The final processing configuration of ANDS based on this analysis is:

- #4 Spline filter with histogram equalization
- Candidate nodule detection
- Sparse convolution smoothing
- Vote accumulation.

3.3 - Tuning the Parameters of ANDS

The optimal processing configuration, #4, was evaluated at 4 radius values (8, 10, 12, and 20; which correspond to nodules from .75 to 2.0 centimeters in diameter), 4 knot spacing values (20, 40, 60, and 120), and 3 resolutions (rescaling by a factor of 2 between spline filtered image and CN centers image; original image rescaled by a factor of 2; and no rescaling) on six films (6, 18, 32, 36, 41, and 44) to determine which configuration of parameters produces optimal detection of the nodules present in the films. The three resolution conditions are illustrated in Fig. 3.3.1. The two resolutions which involve rescaling the image were evaluated because I believed that ANDS would take less time to compute with these smaller images and that the results might be acceptable. The distance metric was used to evaluate the performance of these three parameters. The optimal image resolution was first chosen and then an ANOVA was performed on the remaining parameters at this resolution to determine which parameter contributes a statistically significant amount of variation, given the following model:

$$Y_{ijk} = \eta + F_i + R_j + K_k + F_i \times R_j + F_i \times K_k + \epsilon_{ijk}$$

where:

Y_{ijk} = observed mean

η = effects due to overall mean

F_i = effects due to films

K_k = effects due to knot spacing value

R_j = effects due to radius value

$F_i \times R_j$ = effects due to interactions between film and radius

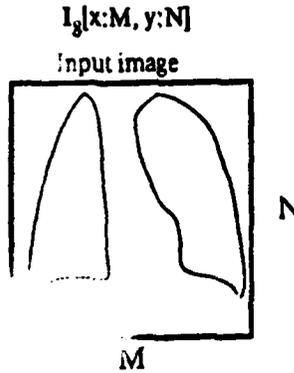
$F_i \times K_k$ = effects due to interactions between film and knot

ϵ_{ijk} = residual effects.

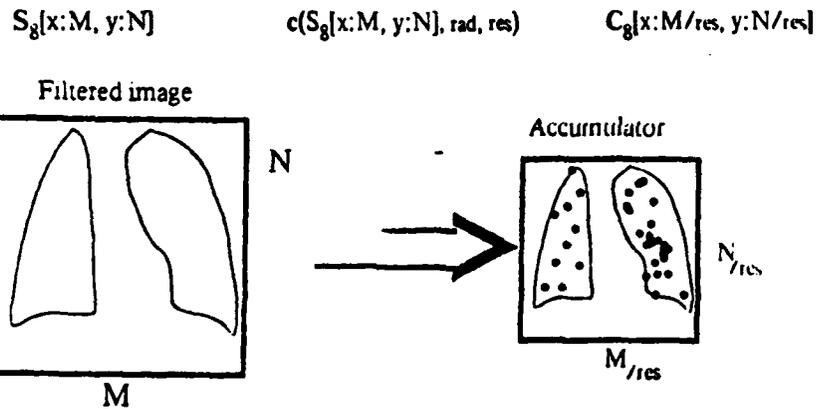
If a parameter makes a statistically significant contribution to the total variance then one might infer that the value of that parameter has an effect on detection. Furthermore, one of the values of the significant parameter might result in a more optimal detection than the other values.

The values of the parameters of ANDS are set at the optimal values of the statistically significant parameters, which were determined by this ANOVA.

Resolution 1: No rescaling.



Resolution 2: Accumulator rescaling. Computed image center coordinates are rescaled before entry into rescaled accumulator image.



Resolution 3: Original image is rescaled by a factor of 2.

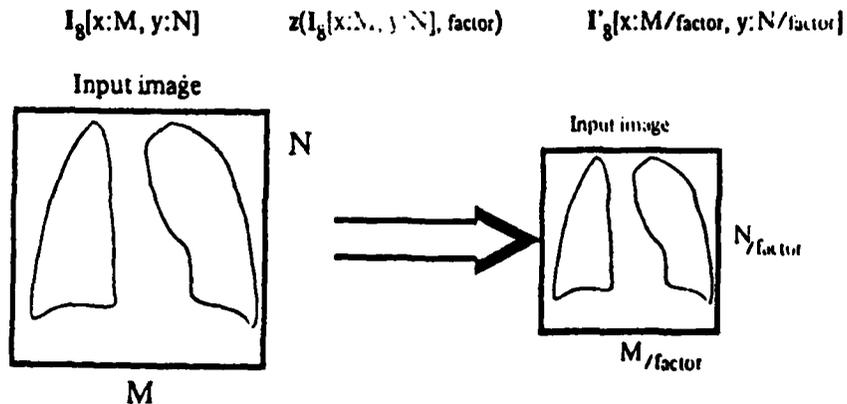


Figure 3.3.1 - The three image resolutions that were evaluated during parameter tuning.

3.3 - Tuning the Parameters of ANDS

Due to limitations imposed on the amount of available computing time and the desire to finish this work in a reasonable amount of time, the scope of the parameter testing had to be limited to 6 films, 4 radius values, 4 knot spacing values, and 3 resolutions (a total of 288 runs). Thus, the optimal parameter values that are reported here are coarse global estimates. Each run averaged about 1.25 hours of real time (15' for file transfer from RIG, checking file for errors, reorienting image, and rescaling image; 20' for spline filtering; 15' for detection of centers of CNs; 20' for sparse convolution smoothing; and 5' for compilation of the list of accumulator values. The entire parameter test required about 360 hours or 15 days of real time to compute. The actual runtime was significantly longer because of failures or rebootings of the machines in the distributed network.

The resolution that is used in ANDS was chosen by inspection of the 95% confidence intervals of the DM means computed over all films, radius values, and knot spacing values. These confidence intervals are illustrated in Fig. 3.3.2. Here, the resolution with the lowest average DM, original image rescaled by a factor of 2, is statistically distinguishable from the other two resolutions.

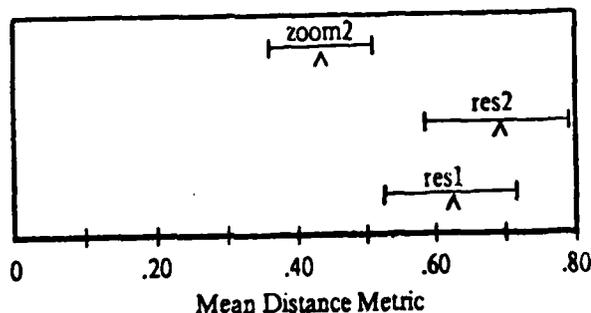


Figure 3.3.2 - 95% confidence intervals on the mean Distance Metrics computed over all films, radius values, and knot spacings for three image resolutions: *res1*, no rescaling of image; *zoom2*, rescaling of input image by a factor of 2; and, *res2*, rescaling between spline filtered image and CN center image by a factor of 2.

An ANOVA was computed over all films, knot spacing values, and radius values for data derived at the chosen resolution. The results of this ANOVA are presented in Table 3.3.1. Effects

due to radius and the radius/film interaction are statistically significant at the $\alpha=0.05$ level. Effects due to films, knot spacing values, and the film/knot interaction are not statistically significant at the $\alpha=0.05$ level. The effects due to knot spacing would be significant at an α -level slightly greater than 0.05.

Source	Sum of Squares	ν	Mean Square	F-ratio
Film	3.6120	5	0.0380	0.96
Radius	3.4504	3	1.1501	29.01*
Knot	0.8602	3	0.2867	7.23
Film \times Radius	2.1724	15	0.1448	3.65*
Film \times Knot	1.2522	15	0.0835	2.11
Error	2.1408	54	0.0396	
TOTAL	13.4881	95	0.1420	

Table 3.3.1 - Results of ANOVA that was performed over all films, which were rescaled by a factor of 2 at all knots and radii. Effects due to radius and film/radius interaction are statistically significant at $\alpha=0.05$ ($P < 0.05$).

Since these analysis failed to show a statistically significant contribution due to knot spacing and the film/knot interaction, a single knot spacing parameter value is incorporated in ANDS. This is the value that results in the lowest DM over all films and radii. The confidence intervals from which this choice was made are given in Fig. 3.3.3. This decision is *ad hoc* because there is no statistically significant difference evident in the confidence intervals. A knot spacing value of 60 was chosen as the parameter of ANDS because this value has the lowest mean, see Fig. 3.3.3.

The significance of radius value and film/radius interaction suggests that no single radius value would suffice in producing optimal detection over all films. Thus, two radius values, 10 and 20, were chosen as the parameters of ANDS. Note: because the chosen resolution is rescaling by a factor of 2, the knot spacing and radius values are reduced by a factor of 2; so, a knot spacing value and radii of 5 and 10 were incorporated into ANDS. Fig. 3.3.4 illustrates the 95% confidence intervals on the DM means these were computed over all radius values and knot spacings. A radius value of 10 pixels (≈ 0.5 cm.) was chosen for the final ANDS because it has the lowest mean DM.

3.3 - Tuning the Parameters of ANDS

although no statistically significant difference is apparent. Although the radius value of 20 pixels (sl.0 cm.) shows a poorer performance (higher DM mean) it was also chosen for incorporation into ANDS because it corresponds to larger nodules.

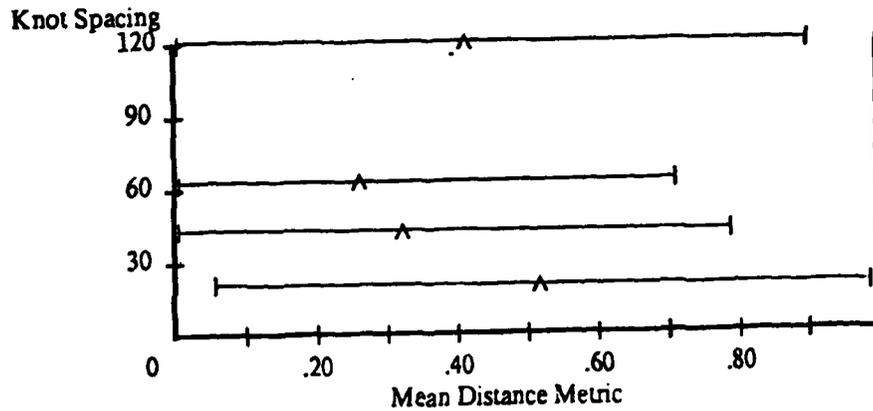


Figure 3.3.3 - 95% confidence intervals on the means of the Distance Metrics for 4 knot spacing values (30, 40, 60, and 120) computed over all films, and radius values.

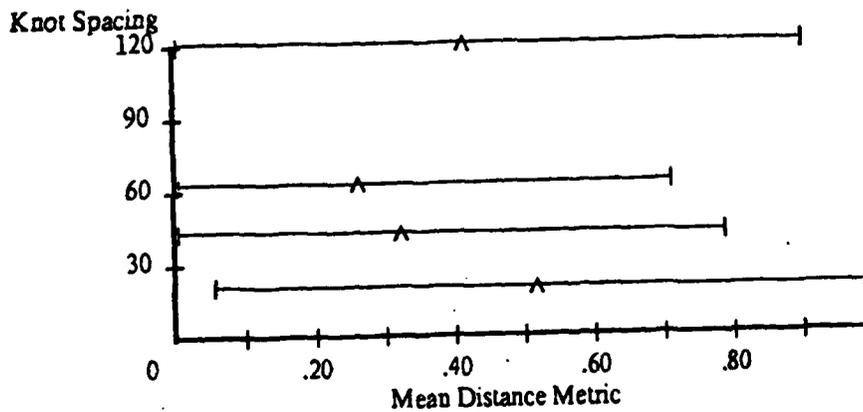


Figure 3.3.4 - 95% confidence intervals on the means of the Distance Metrics for 4 radius values (8, 10, 15, and 20) computed over all films, and knot spacing values.

TR - 120

4 - RESULTS - Evaluation of the Performance of AI in Reducing the False Positive Rate

Given an automated nodule detection system that has been designed to provide optimal detection (over the films and system parameters that were tested) of any pulmonary nodules that are present in a chest radiograph, the final phase of this work is to reduce the number of false positives that are reported by that system. Concomitant with this goal is the mandate of not greatly reducing the number of true positives.

Seven films were omitted from the final test because of errors in digitization (see Fig. 4.1) that prevented computation of the lung borders which are required by the pattern recognizer for determination of the relative location measures. The films that were omitted are numbers 1, 22, 28, 31, and 33. Film 38 was omitted because the nodule was always detected. The nodule is in the lower medial corner of the right lung, see Fig. 4.2. The lung border in the corner of this image act as edges on the border of the nodule. Film 41 was omitted because its image file was accidentally smashed following parameter tuning; time did not permit redigitization. Appendix 9.3 provides the statistics about each film in the ANDS database and summarizes the results of these tests.

Forty-three films were processed at two radii (5 and 10, which correspond to .5 and 1.0 cm, respectively) by ANDS with AI under two conditions and without AI. In the two conditions with AI, the pattern classifier was trained with different numbers of films and tested on the entire database. In the first case, it was trained with 9 films and in the second with 37 films (films that contained nodules whose nodule statistics could be computed).

The nodule appearance statistics from which the features that are used to train the pattern classifier are derived from a smoothed, histogram equalized, windowed region around the CN from the spline filtered image. The top 50 points in the accumulator lists obtained when testing ANDS on the database at two radii were classified (some could not be classified because a 64x64 pixel window around the CN could not be made or because the appearance statistics could not be computed). These classifications were used to train the pattern recognizer. The feature vectors were input to BMDP7M, which runs on the DEC-10.

TR - 120



Figure 4.1 - Image with incomplete scanline, presumably attributable to digitizer. The lung borders could not be computed in images with incomplete scanlines. These images have been omitted from the films evaluated by ANDS with AI because without the relative distance features could not be computed.



Figure 4.2 - Lung #38 was omitted from the evaluation of ANDS because its nodule is always found regardless of parameters and processing configurations. The nodule is in the lower medial corner of the right lung. The lung borders coincide with the margin of the nodule and vote for the nodule

4 - RESULTS - Evaluation of the Performance of AI in Reducing the False Positive Rate

This statistics package computed the weights and constants that are used by the discriminant function. These values were instantiated in the pattern classifier. The input to the Nodule Expert is the feature vector of a CN. The output of the Nodule Expert is a decision - whether or not a nodule was detected. If a CN is recognized as a false positive, it is omitted from the list of CNs that is reported by ANDS. Only CNs that are classified as a nipple or any kind of nodule are kept in the screened list of CNs. The performance of the Nodule Expert is evaluated by subjecting this list of CNs to the performance evaluation procedures that are described in Chapter 2.4. Fig. 4.5 illustrates the result of applying 37-trained ANDS on lung #9 at a radius of 5 pixels.

```

0.5370 ; >>>> DISTANCE Metric for /u/b111/thesis/metric/preA1/191r5k3Uz2p4. <<<<
0.5370 ; >>>> Cum Histo Metric <<<<
25.0000 ; Number of false positives - # of points which are not
; nodules, that lie between the first accumulator point and
; the last detected nodule.
1.0000 ; Percentage of the 2 known nodules which were detected.
50 ; Number of points in accumulator.
0.5000 ; Percentage of positives in 1-th group of 10 accumulator points.
0.0000 ; Percentage of positives in 2-th group of 10 accumulator points.
0.5000 ; Percentage of positives in 3-th group of 10 accumulator points.
0.0000 ; Percentage of positives in 4-th group of 10 accumulator points.
0.0000 ; Percentage of positives in 5-th group of 10 accumulator points.
0. Acc[141, 142] = 9792
1. Acc[132, 200] = 9504
2. Acc[277, 278] = 9120
3. Acc[120, 222] = 8896
4. Acc[353, 117] = 8544
5. Acc[145, 292] = 8064
6. Acc[273, 200] = 7840
7. Acc[283, 250] = 7648
8. Acc[305, 236] = 7648
9. Acc[148, 190] = 7648
10. Acc[291, 285] = 7616
11. Acc[105, 120] = 7328
12. Acc[258, 283] = 7264
13. Acc[155, 288] = 7264
14. Acc[132, 215] = 7232
15. Acc[164, 163] = 7008
16. Acc[369, 221] = 6944
17. Acc[278, 215] = 6880
18. Acc[368, 209] = 6848
19. Acc[129, 164] = 6848
20. Acc[367, 126] = 6816
21. Acc[130, 111] = 6816
22. Acc[147, 165] = 6784
23. Acc[153, 127] = 6688
24. Acc[299, 137] = 6624
25. Acc[188, 332] = 6592
26. Acc[ 69, 113] = 6592
27. Acc[284, 320] = 6560
28. Acc[124, 241] = 6560
29. Acc[163, 224] = 6528
Acc[183, 352] = 6400
Acc[253, 245] = 6400
Acc[173, 246] = 6368
33. Acc[ 58, 176] = 6304
34. Acc[321, 254] = 6240
35. Acc[287, 261] = 6240
36. Acc[ 56, 217] = 6208
37. Acc[304, 177] = 6208
38. Acc[256, 216] = 6176
39. Acc[301, 202] = 6176
40. Acc[317, 187] = 6176
41. Acc[258, 261] = 6144
42. Acc[367, 169] = 6144
43. Acc[292, 151] = 6144
44. Acc[186, 306] = 6080
45. Acc[175, 265] = 6080
46. Acc[ 99, 187] = 6080
47. Acc[371, 81] = 6016
48. Acc[ 56, 157] = 5952
49. Acc[106, 355] = 5920

```

```

0.1667 ; >>>> DISTANCE Metric for /u/b111/thesis/metric/postA12/191r5k3Uz2p4. <<<<
0.1667 ; >>>> Cum Histo Metric <<<<
1.0000 ; Number of false positives - # of points which are not
; nodules, that lie between the first accumulator point and
; the last detected nodule.
1.0000 ; Percentage of the 2 known nodules which were detected.
5 ; Number of points in accumulator.
0. Acc[353, 117] = 8544
1. Acc[105, 120] = 7328
2. Acc[ 69, 113] = 6592
3. Acc[124, 241] = 6560
4. Acc[321, 254] = 6240

```

Figure 4.5 - The result, bottom, of applying AI techniques (37-trained) to reduce the number of false positives in the list of accumulator points produced by the unintelligent ANIDS system, top. The information at the top of each list is a summary of the performance of ANIDS. Data such as the above are compared to assess the effects of the AI techniques on ANIDS.

4 - RESULTS - Evaluation of the Performance of AI in Reducing the False Positive Rate

The performance of the pattern classifier may be visualized in a classification matrix, where the rows represent classes that were taught and the columns represent the classification results of the pattern classifier. Some of the CNs that do not neatly fit into any of the eleven classes were classified as Undetermined (UN) when training the pattern classifier. CNs that are classified as Undetermined were not used when determining the discriminant function. The classification matrix obtained from running ANDS on only the training films when 9-trained is presented in Table 4.1. The classification matrix obtained when running 37-trained ANDS on the training films is presented in Table 4.2. Tables 4.3 and 4.4 present the classification matrices obtained when testing the 9- and 37-trained AND systems on the entire database.

Known class	pct	Classified As:											Count	
		ri	SR	sv	lv	SN	MN	LN	lb	mb	SB	NI		un
RIB	0.53	10	4	1	2	0	1	1	0	0	0	0	0	19
SR	0.43	1	3	0	0	1	1	0	0	0	0	1	0	7
SV	0.73	5	0	62	11	5	1	1	0	0	0	0	0	85
LV	0.67	1	0	7	26	0	0	4	0	1	0	0	0	39
SN	0.50	0	2	2	0	5	1	0	0	0	0	0	0	10
MN	0.56	1	2	0	0	0	5	0	0	0	0	1	0	9
LN	0.80	1	0	0	0	0	0	4	0	0	0	0	0	5
LB	1.00	0	0	0	0	0	0	0	11	0	0	0	0	11
MB	0.88	0	1	2	5	0	0	1	0	63	0	0	0	72
SB	0.33	0	0	0	3	0	1	0	0	0	2	0	0	6
NI	0.50	1	0	1	0	0	0	0	0	0	0	2	0	4
UND	0.00	45	7	25	12	15	3	11	3	2	1	11	0	135

Table 4.1 - Classification matrix for training films for 9-trained ANDS (#s 5, 6, 8, 12, 16, 18, 32, 40, 41) when tested on training films. The abbreviations for the classifications are defined in 2.5.0. 407 CNs were evaluated.

The performances of the Rib and Vascularity Experts are presented in Table 4.5. Here, the row represents all CNs that were detected by these experts and the columns are the class to which the CNs belong (specified when training the pattern classifier). These results were obtained when 9- and 37-trained ANDS were run on the entire database. The performance of the Nodule Expert at both trainings on both the training films and the entire database are presented in Table 4.6. Since several films contain more than one nodule, it is more meaningful (from the patients'

viewpoint) to talk about nodule detection in terms of films which contain nodules that are missed (that is, the nodules that are present are not recognized by the Nodule Expert) rather than absolute percentages of nodules that are missed in a given film. Fig. 4.6 compares the 95% confidence intervals on the average percent correct classification for 9- and 37-trained ANDS when tested on training films and on the entire database.

Known class	pct	Classified As:												Count
		ri	SR	sv	lv	SN	MN	LN	lb	mb	SB	NI	un	
RIB	0.54	90	10	12	7	3	13	8	0	1	4	20	0	168
SR	0.54	1	7	1	0	2	2	0	0	0	0	0	0	13
SV	0.65	14	5	275	72	11	18	1	0	9	15	1	0	421
LV	0.73	1	0	25	136	0	5	4	0	13	2	0	0	186
SN	0.53	1	2	5	0	16	0	0	0	0	1	5	0	30
MN	0.27	3	4	3	8	0	14	4	0	3	3	7	0	51
LN	0.75	0	0	0	2	0	1	9	0	0	0	0	0	12
LB	0.96	1	0	0	0	0	0	0	193	1	0	5	0	200
MB	0.82	2	0	18	31	2	1	0	0	429	36	1	0	520
SB	0.58	0	0	1	1	1	0	0	0	2	7	0	0	12
NI	0.64	0	0	2	0	2	0	0	0	0	0	7	0	11
UND	0.00	158	39	183	30	76	95	22	0	7	29	66	0	705

Table 4.2 - Classification matrix for training films for 37-trained ANDS (trained on all films that contain any nodule(s)) when tested on training films. 2329 CNs were evaluated.

Known class	pct	Classified As:												Count
		ri	SR	sv	lv	SN	MN	LN	lb	mb	SB	NI	un	
RIB	0.50	106	34	18	9	14	1	8	1	1	0	21	0	213
SR	0.29	3	4	0	1	2	1	0	0	0	0	3	0	14
SV	0.68	30	0	332	98	9	2	6	0	13	0	1	0	491
LV	0.67	5	0	37	151	0	0	16	0	17	0	0	0	226
SN	0.57	0	3	6	0	17	3	0	0	0	0	0	0	30
MN	0.18	12	7	5	8	1	9	1	0	2	1	5	0	51
LN	0.33	4	1	0	1	0	2	4	0	0	0	0	0	12
LB	0.97	1	2	0	0	1	0	0	241	0	1	2	0	248
MB	0.79	0	1	42	51	1	0	6	7	475	18	0	0	601
SB	0.25	0	0	3	2	0	1	0	0	2	3	0	0	12
NI	0.23	2	5	1	0	2	0	0	0	0	0	3	0	13
UND	0.00	246	62	187	60	133	10	30	7	12	4	88	0	839

Table 4.3 - Classification matrix for 9-trained ANDS (5, 6, 8, 12, 16, 18, 32, 36, 44) when tested on entire ANDS database. 2750 CNs were evaluated.

4 - RESULTS - Evaluation of the Performance of AI in Reducing the False Positive Rate

Known class	pct	Classified As:													Count
		ri	SR	sv	lv	SN	MN	LN	lb	mb	SB	NI	un		
RIB	0.54	116	11	20	7	5	17	8	0	2	4	23	0	213	
SR	0.50	1	7	1	0	3	2	0	0	0	0	0	0	14	
SV	0.64	14	5	316	96	11	19	1	0	11	17	1	0	491	
LV	0.74	1	0	30	168	0	6	4	0	16	3	0	0	227	
SN	0.53	1	2	5	0	16	0	0	0	0	1	5	0	30	
MN	0.27	3	4	3	8	2	14	4	0	3	3	7	0	51	
LN	0.82	0	0	0	1	0	1	9	0	0	0	0	0	11	
LB	0.96	2	0	0	0	0	0	0	239	1	0	6	0	248	
MB	0.82	3	0	22	34	4	1	0	0	493	42	2	0	601	
SB	0.58	0	0	1	1	1	0	0	0	2	70	0	0	12	
NI	0.54	0	1	2	0	2	0	1	0	0	0	7	0	13	
UND	0.00	187	48	225	34	92	104	23	0	10	35	81	0	839	

Table 4.4 - Classification matrix for 37-trained ANDS (trained on all films that contain any module(s)) when tested on entire ANDS database. 2750 CNs were evaluated.

Performance of Rib Expert on Entire Database

Rib Expert	pct cor't	Taught As:													Count
		ri	SR	sv	lv	SN	MN	LN	lb	mb	SB	NI	un		
RIB	0.15	47	1	27	15	4	3	3	33	87	3	1	87	311	

213 ribs were taught
311 CNs were classified as rib by the Rib Expert

Performance of Vascularity Expert on Entire Database

Vasc Expert	pct cor't	Taught As:													Count
		ri	SR	sv	lv	SN	MN	LN	lb	mb	SB	NI	un		
VASC	0.46	16	0	73	32	0	0	1	0	71	0	0	35	228	

717 vascular structures were taught
228 CNs were classified as vascularity by the Vascularity Expert

Performance of Vision Experts when Tested on Entire Database

Expert	Sensitivity	True Positive Rate
Vascularity	0.15	0.47
Rib	0.22	0.15

Table 4.5 - Performance of vision experts when tested on entire database. The matrices illustrate classifications to which the CNs detected by the experts belong. Sensitivity is the fraction of CNs that belong to the class that was detected by the expert that in fact belong to the correct class.

Performance of Nodule Expert on Entire Database

Training	Sensitivity	True Positive Rate
9-trained	0.58	0.14
37-trained	0.76	0.15

Table 4.6 - Performance of Nodule Expert when tested on entire database at both trainings. This is the performance of the pattern classifier and the classification rule. These results represent the detection performance of ANDS on all nodules in the films that comprise the database. That is these values reflect the ability of ANDS to detect all the nodules in the database.

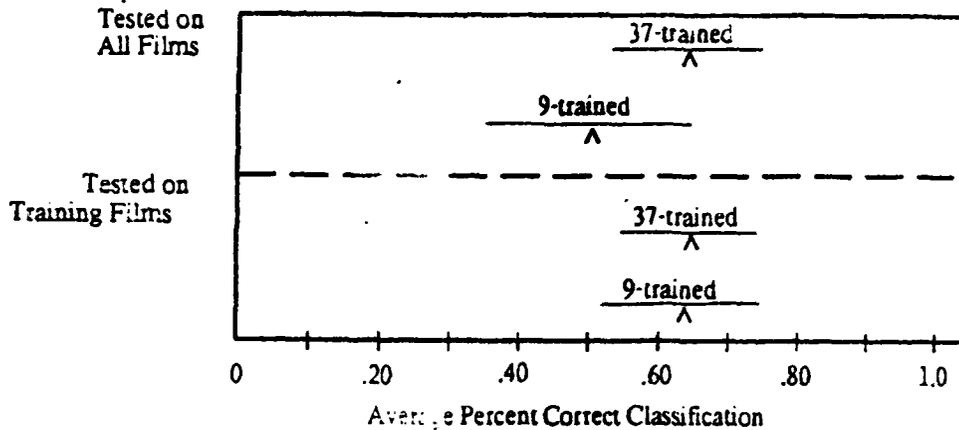


Figure 4.6 - The 95% confidence intervals on the average percent correct classification of ANDS when 9- and 37-trained both when tested on training films and on the entire database

Table 4.7 compares the changes in DM, false positive and true positive between both trainings (9 and 37) and the naive (no AI) ANDS. The true positive rate reported is the average of the true positive rates for each film, where the true positive rate is the fraction of known nodules that is detected. False positive represents the number of non-nodules that lie between the first accumulator point and the last detected nodule. The value reported is the average over all films. The average DM over all films is also reported. Only processing configurations (films and radius) at which any nodule was detected in both the pre-training and the trained systems are included in these analyses. Student's t-statistic is computed for the three metrics as are P-values.

4 - RESULTS - Evaluation of the Performance of AI in Reducing the False Positive Rate

Comparison of the Performance of 9-Trained and naive ANDS

ANDS	DM	METRIC	
		# of False Positive	True Positive Rate
9-trained	0.353	0.588	0.755
naive	0.550	12.235	0.827
Difference	-0.197	-11.647	-0.072
t-statistic	-3.96*	-5.20*	-3.00*
P-Value	$P < 0.0005$	$P \ll 0.0005$	$0.0005 < P < 0.005$

Comparison of the Performance of 37-Trained and naive ANDS

ANDS	DM	METRIC	
		# of False Positive	True Positive Rate
37-trained	0.462	1.809	0.787
naive	0.594	11.915	0.875
Difference	-0.132	-10.106	-0.088
t-statistic	-2.87*	-5.67*	-2.98*
P-Value	$.05 < P < .01$	$P \ll 0.0005$	$0.0005 < P < 0.005$

Table 4.7 - Comparisons of the performance of 9- and 37-trained ANDS with the naive ANDS. These comparisons are made over ANDS configurations (film and radius) which were detected in both naive and trained systems. That is, the differently trained ANDS that are evaluated here were tested on all 37 films in the database. However, it only makes sense to compare films that contain nodules that were detected by the naive system if one is to assess improvement/changes in the detection metrics. Since each film was processed at two radii (there are 37 films in which a nodule was detected) there are 74 possible film/radius combinations that could be compared in the above analyses. Only film/radius combinations in which a nodule was detected by the trained system are included in the comparisons. Thirty-four film/radius combinations (26 different films) were evaluated by the 9-trained ANDS and 47 (32 different films) by the 37-trained ANDS; the reported metrics are means over these numbers of film/radius combinations.

Since it is clinically more important not to miss a radiograph that contains a nodule than to recognize every nodule in a film, Table 4.8 summarizes the false negative rates (in terms of films that were missed) of ANDS at the three trainings.

False Negative Rates For Films

ANDS Training	Single Nodules	Multiple Nodules ¹	All Films
Naive ²	0.12	0.00	0.08
9-trained	0.35	0.18	0.30
37-trained	0.19	0.00	0.14

1 - The average number of nodules per film is 7.5, $s = 11.4$.

2 - The top 50 points in the accumulator list are considered.

Table 4.8 - False negative rates at three trainings. The false negative rates are summarized at the three trainings over all films with one nodule, all films with more than one nodule, and all films with any nodule. Twenty-six films contain a single nodule; 11 films contain multiple nodules. The false negative rates are summarized over all films with one nodule, all films with more than one nodule, and all films with any nodule.

Table 4.9 summarizes the films that were missed by the naive ANDS and which were consequently missed by the trained systems. Table 4.10 summarizes the misclassifications of the CNs in films that were misdiagnosed (that is, no nodule was detected) by ANDS at both trainings.

Summary of Films That Were Misdiagnosed by Naive ANDS

FILM #	Description of Nodule
24	faint shadow of a button on lateral border of right lung
42	pseudo-nodule; near bottom apex of right lung; elongated vertically; well defined margins but non-uniform interior density
43	granuloma; nodule is fuzzy; is on bottom medial border of left lung; poorly defined margin; non-uniform interior density

Table 4.9 - The solitary abnormalities in these three films were not detected by the untrained ANDS at radii of 5 or 10 (pixels). Since the nodules in these films were missed by the untrained system, they were also missed by the trained systems because the trained systems use the accumulator list that is output by the naive system as their input. If a nodule is not anywhere in the list of accumulator points that is produced by the naive system it cannot be detected by a trained system.

4 - RESULTS - Evaluation of the Performance of AI in Reducing the False Positive Rate

Summary of Films That Were Misdiagnosed by 9-trained ANDS

Film #; Radius	Comment on Classification
2; 5	taught as medium nodule, classified as rib; nodule is fibrous and on rib
6; 5	taught as nodule on medial border, classified as large vascularity; nodule is well defined; interior mass is relatively uniform and noticeably brighter than exterior
7; 10	taught as medium nodule, classified as large vascularity; nodule margin is well defined and interior brightness is uniform; nodule is near medial border above the medial midpoint; diameter is greater than 2.0 cm.
20; 5	taught as medium nodule, classified as rib; >2 cm diameter nodule is occluded by clavicle
35; 5	taught as medium nodule, classified as rib; nodule is occluded by rib; margin is fuzzy, interior is uniform
39; 5	two nodules are present in film; [157, 241] taught as medium nodule, classified as small vascularity; well defined margin; near medial and bottom borders of right lung; [371, 296] taught as medium nodule, classified as rib; fuzzy, darker somewhat horizontally elongated in left lung
39; 10	two nodules are present in film; [155, 246] taught as medium nodule, classified as small vascularity; (see above) [372, 289] taught as medium nodule, not classified because of error in computing nodule feature statistics
40; 5	two nodules are present in film; [132, 231] taught as small nodule on border, classified as small vascularity; [367, 277] not taught - unable to compute nodule statistics
40; 10	[132, 231] taught as small nodule on border, classified as small vascularity; [367, 277] not taught - unable to compute nodule statistics

Summary of Films That Were Misdiagnosed by 9-trained ANDS (continued)

Film #; Radius	Comment on Classification
44; 5	taught as nipple, classified as small vascularity; nipple is near mediastinum
44; 10	taught as nipple, classified as rib; no overlapping rib

Table 4.10 - Summary of the films that were missed by the 9-trained ANDS. The above films in addition to those in Table 4.9 were missed by 9-trained ANDS. The above film/radius combinations were detected by the untrained ANDS. The right column explains why the nodule(s) in these films was/were missed. Eleven films were missed by 9-trained ANDS (of 37 films that contain nodules).

Summary of Films That Were Misdiagnosed by 37-trained ANDS

Film #; Radius	Comment on Classification
7; 10	taught as medium nodule, classified as large vascularity
44; 5, 10	taught as nipple, classified as small vascularity

Table 4.11 - Summary of the films that were missed by the 37-trained ANDS. The above films in addition to those in Table 4.9 were missed by 37-trained ANDS. The above film/radius combinations were detected by the untrained ANDS. The right column explains why the nodule(s) in these films was missed. Five films were missed by 37-trained ANDS (of 37 films that contain nodules).

Table 4.12 illustrates that number of CNs from the top of the list of reported candidate nodules that a radiologist must inspect before being 95% or 99% confident of having read a nodule that was detected by ANDS, if one is present in the film. These values are the upper limits of the respective confidence limits on the means that are presented in Table 4.7. The confidence levels are presented for naive, 9-trained, and 37-trained ANDS. The values for the naive system are based on the results for 47 films in which nodules were detected by 37-trained ANDS.

4 - RESULTS - Evaluation of the Performance of AI in Reducing the False Positive Rate

ANDS Training	Confidence Level		True Positive Rate All Films
	95%	99%	
Naive	11	12	.92
9-trained	2	2	.70
37-trained	3	4	.86

Table 4.12 - The number of CN sites from the top of the list of CNs which a radiologist must inspect in order to be 95% or 99% confidence of having read a nodule. The 9-trained limits are based a mean that is obtained from 34 film/radius combination and the 37-trained limits derive from a mean that is obtained from 47 film/radius combinations. The true positive rate is the percentage of all films with at least one nodule that was correctly diagnosed by ANDS.

TR - 120

5 - DISCUSSION

Panoramic View

As a result of this work, an Automated Nodule Detection System, based on a system presented by Ballard [Ballard, 1974], was designed, tuned, and tested on a database of 43 chest radiographs. The reproduction of the original radiographs to digital images was carefully determined and controlled so that a linear transfer was obtained. The final ANDS design was chosen from four candidate system configurations. An ANOVA failed to find any significant difference in nodule detection ability between the tested configurations, when evaluated with six films, this failure was primarily due to a large film-to-film variation which masked any difference due to processing configuration. The configuration that was implemented was chosen primarily because it requires one less step (and is consequently faster) than the top-performing configuration and because its performance measure is within 4% of that configuration. The parameters of ANDS (knot spacing, radius, and image resolution) that result in optimal detection of nodules in five films were determined when tuning ANDS. The knot spacing value, the parameter of the spline filter, was found to have no statistically significant effect on the detection of the nodules of various sizes in the five films that were evaluated. A knot spacing of 60 pixels was chosen because this results in the highest mean detection performance over all five films tested; this value is not statistically significant at ($\alpha=0.05$). The amount of variation contributed by the radius value that is used by the CN Expert was shown to be statistically significant ($\alpha=0.05$) and to have a statistically significant interaction with film (that is, nodule size). Two radius values were chosen and are implemented in ANDS. These radius values are 5 and 10 pixels. A radius value of 5 pixels was chosen because the detector performance on all five films was the highest at this value, although the performance at this value is not statistically different from those of the other tested values. A radius value of 10 pixels was chosen because it corresponds to a 2 centimeter (diameter) nodule. The image resolution that was chosen for incorporation in ANDS is: rescaling of the original image by a factor of 2. That is, each dimension of the original digital image is reduced by half. A reduction in high frequency image noise may be the cause of the improved detection of nodules in half-size images. This noise could also have been reduced by averaging multiple scanings of the image. Presumably, the noise is random and is a result of digitization.

Performance of the Experts

The performance of this ANDS was assessed on the entire database. A Nodule Expert (a pattern classifier with a set of classification rules) was trained twice, first with 9 films and then with 37 films (all of these films contained at least one nodule). The Nodule Expert detected 76% of all known

nodules in the entire database when 37-trained and detected only 58% when 9 trained. This difference is presumably due to the more extensive training. The contribution of the Rib Expert may be judged significant because its output value (whether or not a rib was detected) is used by the pattern recognizer. The output of the Vascularity Expert, however, is not used by the pattern recognizer. Presumably, this may be attributed to the strength of the *features* that are used by the pattern classifier, rather than to a weakness of the Expert. That is, the features may provide a more concise measure/description of vascularity than does the Expert. The Vascularity Expert essentially adds knowledge about linear clustering of CNs; this may not be necessary to recognize vascularity. Knowledge about the CN's appearance and relative location in the lung is perhaps more pertinent.

Comparison of various trainings

Initially, I had planned to compare only the naive ANDS and 9-trained ANDS. Since 11 films (of 37 films that contain at least one nodule) were missed by the 9-trained system, the 37-trained system was developed and tested. Five films were missed by the 37-trained ANDS. That is, three films in addition to the two films which were missed by the naive system were missed by 37-trained ANDS. Both trained systems proved capable of reducing the number of false positives (Table 4.7) at a statistically significant level ($P \ll 0.0005$). As a consequence of reducing the false positive rate, the average DM for all films tested by both trained systems was also significantly reduced; this is desirable. However, a statistically significant ($P < 0.005$) decrease in the true positive rate (the fraction of all known nodules that are detected) also accompanied the reduction in false positives; this may be *tolerable*, although not entirely *desirable*. The true positive rate (in terms of nodules) for the naive system, 0.875, is reduced to 0.787 for 37-trained ANDS; this is an 11% reduction. The true positive rate for films decreased only 6.5% from 92% (for the naive system) to 86% (for 37-trained ANDS). This suggests that more nodules in films containing multiple nodules are being missed by the 37-trained system. That is, it seems that the discrepancy in the true positive rates is not due to missed nodules in films that contain only single nodules.

The trade-off between true and false positives

One is faced with a trade-off when one desires to reduce the number of false positives (or the number of CN sites that the radiologist must inspect); this trade-off is between the number of false positives and the true positive rate. When one desires fewer false positives one must consequently accept the possibility of detecting fewer nodules. Of course the detection rates of the system may be improved by further training, but how much improvement can be gained and how much training would be required is not known. A system with lower false positive and higher true positive rates may be possible.

6 - Conclusion

Pattern recognition techniques and procedurally driven image experts are capable of reducing the number of CN sites that a radiologist must inspect from at most 12 to at most 4 in order to be 99% confident of having inspected any nodule(s) detected by 37-trained ANDS. The radiologist must be willing to accept a film true positive rate of 88% (as opposed to a film true positive rate of 92%) for the convenience of having fewer points to inspect. These film true positive rates are derived from 37 films which contain nodules that were evaluated by ANDS.

TR - 120

AD-A150 940

DESIGN TUNING AND PERFORMANCE EVALUATION OF AN
AUTOMATED PULMONARY NODULE. (U) ROCHESTER UNIV NY DEPT
OF COMPUTER SCIENCE W LAMPETER FEB 83 TR-120

2/2

UNCLASSIFIED

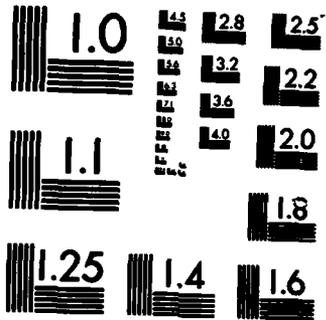
N00014-80-C-0197

F/G 6/5

NL



END
FILMED
DTR



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

7 - Future Work

Train ANDS with more films and evaluate results

Table 4.9 illustrates the effect of training ANDS with 9 and with 37 films. The film false negative rate decreases from 30% to 14% when ANDS is trained with more films. I believe that ANDS can be made more effective if it is trained with more (about 100) films.

Implement parts of ANDS in VLSI hardware

The spline filter, histogram equalization, circle detector, convolution, image search, feature computation, and pattern recognition phases of ANDS may be implemented in hardware for added speed of execution.

Compare ANDS with radiologists

Time has not allowed the completion of a comparison between radiologists, who are instructed to find all nodules in a subset of ANDS films. The radiologists are also instructed to rate their confidence that each is a nodule. These results will be reported at a later time.

TR - 120

8 - References

- American Medical News*, p.9, Oct. 29, 1973.
- Baker, K.D., D.Phil, G.D. Sullivan, "Multiple bandpass filters in image processing" *IEE Proc.* 127-E:5, pp. 173-184, Sept. 1980.
- Ballard, D.H., H.W. Fischer, "Computer assisted detection of lung nodules," pp. 454-457.
- Ballard, D.H., J. Sklansky. "A ladder-structured decision tree for recognizing tumors on chest radiographs" *IEEE Trans. on Comp. C-25:5*, pp. 503-513, May 1976.
- Ballard, D.H., J. Sklansky, "Tumor detection in radiographs," *Computers and Biomedical Research*, 6, pp. 299-321, 1973.
- Bateson, E.M., "The solitary circumscribed bronchogenic carcinoma," *Brit. J. Radiol.*, 37, pp. 598-607, 1964.
- Bateson, E.M., "An analysis of 155 solitary lung lesions illustrating the differential diagnosis of mixed tumors of the lung," *Clin. Radiol.*, 16:1, pp. 51-65, Jan. 1965.
- Brigham, E.O. *The Fast Fourier Transform*. Prentice-Hall Inc. New Jersey, 1974.
- Carmody, D.P., C.F. Nodine, H.L. Kundel, "Global and segmented search for lung nodule of different gradients," *Inv. Radiol.* 15:3, pp.224-233, May-June 1980.
- Cornsweet, T.N. *Visual Perception*. Academic Press, New York, pp. 324-330, 1970.
- Davis, E.W., J.W. Peabody Jr., S. Katz, "The Solitary Pulmonary Nodule," *J. Thor. Surg.*, 32:6, Dec. 1956.
- Dixon, W.J., M.B. Brown. *BMDP-79*. University of California Press, Berkeley. 1979.
- Duda, R.O., P.E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *CACM*, 15:1, pp. 11-15, Jan. 1972.
- Duda, R.O., P.E. Hart. *Pattern Classification and Scene Analysis*. Wiley Interscience. John Wiley & Sons, New York. 1973.
- Garland, H.L., "Studies on the accuracy of diagnostic procedures," *Am. J. Roent.*, 82:1, pp. 25-38, July 1959.
- Goldmeier, E., "Limits of visibility of bronchogenic carcinoma," *Am. Rev. of Resp. Dis.*, 91:2, pp. 232-239, Feb. 1965.
- Gonzales, R.C., P.Wintz. *Digital Image Processing*. Addison-Wesley Publishing, Reading Massachusetts, 1977.
- Good, C.A., T.W. Wilson, "The solitary circumscribed pulmonary nodule," *JAMA*, 166:3, pp. 210-215, Jan. 18, 1958.
- Good, C.A., R.T. Hood, J.R. McDonald, "Significance of a solitary mass in the lung," *Amer. J. Roentgenol.*, 70:4, pp. 543-554, Oct. 1953.

- Hall, D.L., et al., "Direct computer diagnosis of rheumatic heart disease," *Radiol.* 101, pp. 497-509, Dec. 1971.
- Hall, E.L., B.K. Rouge, R.P. Kruger, "Automated chest x-ray analysis," *SPIE*, 89, pp. 109-118, 1976.
- Hemmingsson, A., B. Jung, T. Loennerholm, "Perception of simulated lesions in the lung," *Acta Radiol.*, 16, pp. 494-502, 1975.
- Henschke, C.I., S.J. Hessel, B.J. McNeil, "Automated diagnosis in radiology," *Inv. Radiol.*, 14:3, pp. 195-201, May-June 1979.
- Holin, S.M., et al., "Solitary pulmonary nodules found in a community-wide chest roentgenographic survey," *Amer. Rev. Tuberc. Pul. Dis.*, 79:4, pp. 427-439, April 1959.
- Jagoe, J.R., Paton, K.A., "Reading chest radiographs for pneumoconiosis by computer," *Brit. J. Ind. Med.*, 32, pp. 267-272, 1975.
- James, T.H. ed., *The Theory of the Photographic Process*. Macmillan Publishing, New York, 1977.
- Jarvis, C.L., D.W.G. Moore, "Quantitative analysis of a class of biomedical images by an image processing system," *Computers and Biomedical Research*, 5, pp. 540-560, 1972.
- Kelsey, C.A., et al., "Effect of size and position on chest lesion detection," *Am. J. Roentgenol.*, 129, pp. 205-208, Aug. 1977.
- Kruger, R.P., E.L. Hall, A.F. Turner, "Hybrid optical-digital radiography-based system for lung disease detection," *Appl. Opt.*, 16:10, pp. 2637-2646, Oct. 1977.
- Kruger, R.P., E.L. Hall, S.J. Dwyer, G.S. Lodwick, "Image enhancement and feature extraction for radiographic images."
- Kundel, H.L., D.J. Wright, "The influence of prior knowledge on visual search strategies during the viewing of chest radiographs," *Radiol.*, 93, pp. 315-320, Aug. 1969.
- Kundel, H.L., G. Revesz, H.M. Stauffer, "Evaluation of a television image processing system," *Inv. Radiol.*, 3, pp. 44-50, 1968.
- Kundel, H.L., G. Revesz, H.N. Stauffer, "The Electro-Optical Processing of Radiograph Images," *Rad. Clin. of N. Amer.*, 8:3, pp. 447-460, Dec. 1969.
- Kundel, H.L., G. Revesz, M.C. Ziskin, F.J. Shea, "The image and its influence on quantitative radiological data," *Inv. Radiol.*, 7:4, pp. 187-205, July-August 1972.
- Kundel, H.L., "Visual sampling and estimates of the location of information on chest films," *Inv. Radiol.* 9:2, pp. 87-93, March-April 1974.
- Kundel, H.L., G. Revesz, L. Toto, "Contrast gradient and the-detection of lung nodules," *Invest. Radiol.*, 14, pp. 18-22, Jan.-Feb. 1979.
- Kundel, H.L., "Peripheral vision, structured noise and film reader error," *Radiol.*, 114, pp. 269-273, Feb. 1975.
- Kundel, H.L., C.F. Nodine, D. Carmody, "Visual scanning, pattern recognition and decision-making in pulmonary nodule detection," *Invest. Radiol.*, 13, pp. 175-181, May-June 1978.

8 - References

- Kundel, H.L., G. Revesz, "Lesion conspicuity, structured noise, and film reader error." *Am. J. Roentgenol.*, 126, pp. 1233-1238, 1976.
- Lodwick, G.S. T.E. Keats, J.P. Dorst, "The coding of roentgen images for computer analysis as applied to lung cancer," *Radiol.* 81:2, pp. 185-200, Aug. 1963.
- McFarland, W.D., "Automated chest x-ray analysis," *SPIE*, 206, pp. 146-150, 1979.
- Mackworth, N.H., "Visual noise causes tunnel vision," *Psychon. Sci.*, 3, pp. 67-68, 1965.
- Maffei, L., "Spatial frequency channels and neural mechanisms," in *Handbook of Sensory Physiology*, vol. 8, "Perception," (edited by Held, Liebowitz, and Teubert), Springer Verlag, Berlin, 1978, pp. 37-66.
- Meyers, P.H., et al., "Evaluation of a computer-retrieved radiographic image," *Radiol.*, 81, pp. 201-206, Aug. 1963.
- Moore, R., R.S. Ledley, H.C. Sing. "Application of Automatic Processing of Methods to the Radiographic Image," *Rad. Clin. of N. Amer.*, 8:3, pp. 473-483, Dec. 1969.
- Movshon, J.A., P. Lennie, "Pattern selective adaptation in visual cortical neurones," *Nature*, 278:5707, pp. 850-852, Apr. 26, 1979.
- Nevatia, R. "Locating object boundaries in textured environments," *IEEE Trans. on Comp.*, pp. 170-175, Nov. 1975.
- Nilsson, A.J., *Problem Solving Methods in Artificial Intelligence.*, McGraw-Hill Co., 1971.
- Optronics International Inc., *Operation and Maintenance Manual 10x10 Colorscan System C-4100.* Optronics International, Chelmsford Massachusetts, no date.
- Pearson, J.J., et al., "Automatic inspection of artillery shell radiographs," *SPIE*, 155, pp. 214-221, 1978.
- Preston, K., M. Onoe. *Digital Processing of Biomedical Images*, Plenum Press, New York, 1976.
- Revesz, G., H.L. Kundel, M.A. Graber, "The influence of structured noise on the detection of radiologic abnormalities," *Inv. Radiol.*, 9:6, pp. 479-486, Nov.-Dec. 1974.
- Revesz, G., H.L. Kundel, "Psychophysical studies of detection errors in chest radiology," *Radiol.*, 123, pp. 559-562, June 1977.
- Ryan, T.A., B.L. Joiner, B.F. Ryan. *Minitab - Student Handbook*. Duxbury Press, Mass., 1976.
- Selzer, R.H., "Computer Processing of the Roentgen Image," *Rad. Clin. of N. Amer.*, 8:3, pp. 461-471, Dec. 1969.
- Seybold, W.D., "Solitary or 'coin' lesions of the lung," *Postgrad. Med.*, 36, pp. 424-430, 1964.
- Shea, F.J., M.C. Ziskin, "Visual system transfer function and optimal viewing distance for radiologists," *Inv. Radiol.*, 7:3, pp. 147-151, May-June 1977.
- Siegelman, S.S., et al., "CT of the solitary pulmonary nodule," *AJK*, 135, pp. 1-13, Jul. 1980.
- Sklansky, J., "On the Hough technique for curve detection," *IEEE Trans. Comput.*, C-27:10, pp.

923-926, Oct. 1978.

- Sklansky, J., "Applications of pattern recognition in radiology," pp. 449a-449f.
- Sklansky, J., L.P. Cordella, S. Leviardi, "Parallel detection of concavities in cellular blobs," *IEEE Trans Comput.*, C-25:2, pp. 187-195, Feb. 1976.
- Steele, J.D., "The solitary pulmonary nodule," *J. Thor. Cardio. Surg.*, 46:1, pp. 21-36, July 1963.
- Stutik, F.P., Tockman, M.S., "Radiographic screening in the early detection of lung cancer," *Radiol. Clin. N. Amer.*, 16:3, pp. 347-366, 1978.
- Taylor, R.R., L.N. Rivkin, J.M. Salyer, "The solitary pulmonary nodule," *Ann. Surg.*, 147:2, pp. 197-202, Feb. 1958.
- Theros, E.G., "Varying manifestations of solitary pulmonary neoplasms: a radiologic-pathologic correlative study," *AJR*, 128, pp. 116-123, June 1977.
- Tully, R.J., R.W. Conners, C.A. Harlow, G.S. Dawick, "Towards computer analysis of pulmonary infiltration," *Inv Radiol.*, 13, pp. 298-305, July-Aug. 1978.
- Tuddenham, W.J., "The visual physiology of roentgen diagnosis," *Amer. J. Roentgenol.*, 78:1, pp. 116-123, July 1957.
- Turner, A.F., E.L. Hall, R.P. Kruger, "Measurements of lung vascularity for early diagnosis of pulmonary edema," *SPIE*, 89, pp. 122-128, 1976.
- Vivas, J.R., S.F. Crabtree, "The significance of the radiographic density of solitary lesions of the lungs," *Amer. Prac.*, 4:12, pp. 8-16, Dec. 1953.
- Walske, B.E., "The solitary pulmonary nodule," *Dis. of Chest*, 49:3, pp. 302-304, Mar. 1966.
- World Health Organization. Epidemiological and Vital Statistics Report #18. World Health Organization, Geneva, 1965.
- Yamamura, Y., et al., "A diagnostic data-processing of respiratory diseases centering on lung cancer," *Med. J. Osaka Univ.*, pp. 403-415, March 1965.
- Yerushalmy, J., "Reliability of chest radiography diagnosis of pulmonary lesions," *Amer. J. Surg.*, 89, pp. 231-240, Jan. 1955.
- Ziskin, M.C., F.J. Shea, "Linear spatial frequency filtering for enhancement of radiographic images," *Diag. Radiol.*, 103, pp. 101-105, Apr. 1972.
- Ziskin, M.C., F.J. Shea, H.L. Kundel, G. Revesz, "Accuracy of radiological decision making," (source unknown), pp. 41-47.
- Zwerling, H.B., et al., "The clinical importance of lesions undetected in a mass radiographic survey of the chest," *Amer. Rev. of Tuberc.*, 64, pp. 249-255, 1951.

Appendix 9.1 - Calibration of Optronics Scanner

The Optronics C-4100 scanner was calibrated according to the maintenance manual [Optronics, date unknown] using Kodak neutral density filters and the #2 step tablet. Despite this calibration, few steps above 2.00 on the step tablet were evident in the digitized image. A piece of film (Kodak Commercial) with a density of 2.0 and the 2.0 neutral density filter were scanned while the current from the photomultiplier was measured. The measured currents for both films were not the same. This difference is attributed to the transmission characteristics of the objects that were measured. The density of the film is due to *silver filaments* that are suspended in the gelatin matrix, while the density of the ND filter is due to *carbon particles* in a gelatin matrix. This discrepancy is due to a higher Callier coefficient for the film [James, pp. 488-489]. The illumination/collection geometry of the scanner is that of a microdensitometer. The light that is scattered by the larger grains in the film is never collected by the microdensitometer. No light is lost due to scattering by the ND filter; its Callier factor is very near 1.0.

Thus, the scanner was calibrated using the step tablet, whose density characteristics more closely approximate those of the reduced radiograph images that are digitized. The scanner was calibrated to provide the optimum discrimination between two steps (densities about 2.6-2.75). These two steps were digitized and their digital values were compared (in the final image) and potentiometer R52 was adjusted to obtain a maximum difference between these steps.

TR - 120

Appendix 9.2 - Computation of ANOVA: Processing Methods

This ANOVA compares four image processing configurations that were used to detect nodules in five chest radiographs. The model is:

$$Y_{ij} = \eta + F_i + P_j + \epsilon_{ij}$$

where:

- Y_{ij} = observed mean
- η = general mean
- F_i = effects due to films
- P_j = effects due to processing methods
- ϵ_{ij} = effects due to error

Four measures (true positive rate, TP; false positive rate¹, FP; and a histogram metric², QHM) were combined in a weighted average for each film/processing combination to obtain the values that are used in the ANOVA. The weighted average is:

$$\text{metric} = .5TP + .3(1 - FP^1) + .2QHM$$

¹The false positive rate that was used when computing this ANOVA is different from the one described earlier in this thesis. It is defined as the ratio of the number of non-nodules that lie between the first accumulator point and the last detected nodule and the total number of nodules in the accumulator list.

²QHM is a histogram metric that is only used in this ANOVA. It is defined as:

$$QHM = \sum_{q=0}^{Q-1} (1 - (1/Q)q)h[q]$$

where:

- $h[]$ = a histogram with Q entries; each entry represents the percentage of detected nodules that were located in a given Q-ile in the list of CNs
- Q = the number of quantiles, equal-size divisions of the list of CNs

TR - 120

The data from which the ANOVA is computed:

film	Processing Method				F_i
	P1	P2	P3	P4	
6	1.00	1.00	0.99	1.00	3.985
32	0.47	0.52	0.39	0.53	1.912
36	0.99	0.98	0.94	0.99	3.890
18	0.34	0.56	0.50	0.64	2.035
41	0.74	0.91	0.57	0.65	2.864
P_j	3.540	3.958	3.375	3.813	14.686

The results of this ANOVA are summarized in Table 3.2.1.

Description of the ANDS Database
Summary of Nodule Detection Tests

LUNG	PATHOLOGY	RADIUS(cm.)	AGE	SEX	#NODULES	SUCCESSFUL ¹ RADII		
						preAI	postAI	postAI2
1 3	met(colon)	NC	51	F	1			
2 2	met(colon)	0.61	51	F	1	10	-	5
3 1	met(colon)	0.88	51	F	1	B	10	B
4	granuloma	0.83	21	M	1	B	10	B
5	granuloma	0.56	50	M	1	B	B	B
6	granuloma	0.82	69	M	1	B	-	10
7	hamartoma	NC	50	M	1	10	-	-
8	met(?)	1.02	53	M	1	B	10	B
9	nipples	0.36	73	M	2	B	5	5
10	granuloma	0.83	49	M	1	B	10	B
11	nipples	0.72	65	F	2	B	B	B
12	hamartoma	0.97	44	F	1	10	10	10
13	nipples	0.93	56	M	1	10	10	10
14	Hodgkins	0.82	36	M	1	B	10	B
15	granuloma	0.70	29	M	1	B	10	5
16	nipples	0.44	55	F	2	B	B	B
17 2	met(colon)	0.40	54	M	1	5	5	5
18 1	met(colon)	0.62	56	M	4	B	B	B
19	carcinoma	0.65	71	M	1	10	10	10
20	carcinoma	NC	51	M	1	B	-	B
21	granuloma	0.44	53	F	1 ⁺	B	5	10
22	granuloma	0.49	56	M	1	NC		
23	granuloma	0.83	63	F	1	B	B	B
24	button	NC	44	M	1	-	-	-
25	nipples	0.56	76	M	2	B	B	B
26	met(kidney)	0.68	50	M	1	5	5	5
27	met(melanoma)	0.58	21	F	22	B	B	B
28	granuloma	0.83	50	F	1	NC		
29	carcinoma	NC	69	F	1	B	10	B
30	carcinoma	0.88	55	F	1	10	10	10
31	carcinoma	0.61	52	F	1	B	NC	
32	met(salivary)	0.53	64	F	37	B	B	B
33	met(breast)	0.94	39	F	1		NC	
34	met(breast)	0.61	54	F	5	B	5	B
35	carcinoma	0.99	64	M	1	B	-	B
36	carcinoma	0.69	63	M	1	B	5	5
37	hamartoma	NC	64	M	1	B	5	5
38	carcinoma	1.35	57	F	1	NC		
39 1	pseudo-nodule	0.61	51	M	2	B	-	5
40 2	pseudo-nodule	0.48	53	M	2	B	-	5
41 3	pseudo-nodule	0.48	55	M	2	B	NC	
42	pseudo-nodule	0.42	48	M	1	-	-	-
43	granuloma	1.27	45	F	1	-	-	-
44	nipple	0.49	67	F	1	B	-	-
45	none		29	F				
46	none		35	F				
47	none		64	M				
48	none		77	M				
49	none		30	M				
50	none		56	M				

No results were obtainable from entries in boldface because of digitization errors (dropped scanlines) or smudged images

files. Films #38 was omitted because nodule is always found because it is in lower medial corner of right lung.

¹ Successful radii = radii at which at least one of the nodules present in the films was detected; H=both radius 5 and 10 (pixels); 5=5 pixels (.5 cm); 10=10 pixels (1.0 cm). All films were processed by ANDS at two radii 5 and 10 pixels.

NC = Not Computed: the data were not computed because of error in computing nodule statistics (radius data) or because of scanline error (could not compute lung boundaries).

|n = Adjacent film numbers that are followed by a | are part of a series. The number, n, after the | indicates the chronological position of that film in the series.

END

FILMED

4-85

DTIC