

Computer-Aided Diagnosis in Chest Radiography

Bram van Ginneken

Chest radiographs account for more than half of all radiological examinations; the chest is the "mirror of health and disease". This thesis is about techniques for computer analysis of chest radiographs. It describes methods for texture analysis and segmenting the lung fields and rib cage in a chest film. It includes a description of an automatic system for detecting regions with abnormal texture, that is applied to a database of images from a tuberculosis screening program.

The electronic version of this thesis can be found at
<http://www.isi.uu.nl/~bram/thesis>

The chest radiographs shown on the front and back cover have been taken during a tuberculosis screening program. The order in which they are displayed is determined by the computer algorithm described in Chapter 8 of this book. According to this algorithm, the abnormality increases, column by column, from top-left on the back cover to bottom-right on the front cover.

Computer-Aided Diagnosis in Chest Radiography

Bram van Ginneken

This book was typeset by the author using L^AT_EX2_ε.

Copyright © 2001 by Bram van Ginneken. All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the author.

ISBN 90-393-2654-1

Printed by Ponsen & Looijen, Wageningen.

Computer-Aided Diagnosis in Chest Radiography

Computer-ondersteunde diagnose in
Röntgenbeelden van de thorax
(met een samenvatting in het Nederlands)

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR AAN DE UNIVERSITEIT UTRECHT
OP GEZAG VAN DE RECTOR MAGNIFICUS, PROF. DR. H. O. VOORMA, INGEVOLGE
HET BESLUIT VAN HET COLLEGE VOOR PROMOTIES IN HET OPENBAAR TE
VERDEDIGEN OP VRIJDAG 9 MAART 2001 DES MIDDAGS TE 2:30 UUR

DOOR

Bram van Ginneken

geboren op 11 juni 1970 te Nuenen.

Promotor: **prof.dr.ir. M. A. Viergever**
University Medical Center Utrecht

Co-promotor: **dr.ir. B. M. ter Haar Romeny**
University Medical Center Utrecht

The research described in this thesis was carried out at the Image Sciences Institute, University Medical Center Utrecht (Utrecht, the Netherlands), under the auspices of ImagO, the Utrecht graduate school for Biomedical Image Sciences. The project was financially supported by the Netherlands Ministry of Economic Affairs, within the framework of the Innovation Oriented Research Programme (IOP Beeldverwerking, project number IBV96009).

Financial support for publication of this thesis was kindly provided by Applicare Medical Imaging B.V., Nucletron B.V., Philips Medical Systems Nederland B.V., R2 Technology Inc., and the Röntgen Stichting Utrecht.

Contents

1	Introduction and Outline	13
	Tuberculosis	13
	Chest radiography	15
	Outline of this thesis	17
2	Computer Analysis of Chest Radiographs — A Review	19
	General processing	21
	Enhancement	21
	Subtraction techniques	21
	Segmentation	22
	Lung fields	22
	Rib cage	24
	Other structures	26
	Analysis	27
	Size measurements	27
	Nodule detection	29
	Texture analysis	34
	Other tasks	35
	Conclusions and recommendations	35
3	Segmenting the Lung Fields	39
	Purpose, materials and methods	40
	Approaches to lung segmentation	41
	Matching	41
	Rule-based detection of lung contours	43
	Pixel classification	47
	Hybrid scheme	49
	Algorithms for lung segmentation	49
	Results	52
	Discussion	53
	Conclusions	59

4	Applications of Locally Orderless Images	61
	Locally orderless images	62
	Median and maximum mode evolution	64
	Switching modes in bi-modal histograms	66
	Histogram transformations	66
	Texture segmentation based on local histograms	71
	Concluding remarks	73
5	Texture Classification from Locally Orderless Images	75
	Previous work on texture classification	76
	Theory	77
	Locally orderless images	77
	Filter banks	79
	Joint probability distributions	79
	Feature extraction	80
	Dimensionality and relations between scale parameters	81
	Experiments	83
	Data	83
	Classifier	85
	Texture feature sets	86
	Results and discussion	88
	Conclusions	92
6	Active Shape Model Segmentation with Optimal Features	95
	Active Shape Models	97
	Shape Model	97
	Appearance Model	99
	Multi-resolution framework	100
	Optimization algorithm	100
	Improving ASMs	100
	ASMs with optimal features	101
	Computational considerations	105
	Experiments	106
	Materials	106
	Methods	107
	Results	107
	Discussion and conclusions	110
7	Segmenting the Rib Cage	117
	Relation to previous work	118
	Problem definition	118
	Materials	119
	Rib model	119
	Statistical rib cage model	120
	Rib cage fit measure	124
	Fitting the rib cage	125

Discussion	127
8 Texture Analysis in Chest Radiographs	133
Materials	135
TB database	135
ID database	137
Methods	137
Overview	137
Segmentation	137
Texture analysis	140
Classifying texture feature vectors	140
Classifying images	141
Results	142
Segmentation	142
Texture analysis and classification	143
Discussion	146
Conclusion	152
9 Summary and General Discussion	155
Summary	155
General discussion	158
Bibliography	163
Samenvatting	179
Publications	181
Acknowledgements	183

Prologue

MARTIN PHILIPS is an ambitious radiologist working at the Hobson University Medical Center in New York. He is working together with William Michaels, a brilliant computer scientist who completed his PhD at MIT at the age of nineteen and went on to become the head of a newly created Division of Artificial Intelligence. One morning Michaels visits Philips with a special surprise.

Philips flipped open the case and pulled out the cassette. It wasn't a musical recording. It was a computer program.

"How far did we get?" asked Philips almost in a whisper.

"It's the whole thing," said Michaels.

"No!" said the incredulous radiologist.

"You know the last material you gave me? It worked like a charm. It solved the problem of density and boundary interpretation. This program incorporates everything you've included in all your flow sheets. It will read any skull X ray you give it, provided you put in into that piece of equipment over there." Michaels pointed to the back of Philips' office. There on top of Philips' worktable was a TV-sized electrical apparatus. It was obvious that it was built as a prototype rather than production model.

"I don't believe it," said Philips, afraid that Michaels was teasing him again.

"Neither do we," admitted Michaels. "Everything just suddenly fell together." He walked over and patted the top of the computer unit. "All the work you'd done in breaking down the problem-solving and pattern recognition aspects of radiology not only made it apparent we needed new hardware but also suggested the way to design it. This is it."

"Looks simple from outside."

"As usual, appearances are deceptive," said Michaels. "The innards of this unit are going to revolutionize the computer world."

"And think of what it's going to do to the field of radiology if it can really read X rays," said Martin.

"It will read them," said Michaels, "but there could still be bugs in the program."

Philips feeds film after film into RadRead as Michaels has called his masterpiece, and is stunned by the program's baffling performance.

As soon as the typewriter ceased its rapid staccato, Philips tore off the report. He read it with Denise looking over his shoulder.

“Amazing!” said Philips when he’d finished. “The computer certainly agrees with our impression. And it remembered that it had seen the same density pattern on Lisa Marino’s X ray, and on top of that it asks me to tell it what this density variation is! This thing is goddamn amazing. It wants to learn! It’s so human it scares me.”

Philips grabbed the phone and punched out Michaels’ home number. When Michaels answered, he told him he had not been able to follow up on Marino, but was going to get definitive films in the morning. He added that the computer had actually asked him to tell it what the abnormal density changes were.

“Remember,” said Michaels, “the program approaches radiology the same way you do. It’s your techniques that it utilizes.”

“Yeah, but it’s already better than me. It picked up density variation when I didn’t see it. If it uses my techniques, how do you explain that?”

“Easy. Remember, the computer digitizes the image into a two-hundred-fifty-six by two-hundred-fifty-six grid of pixel points with gray values between zero and two hundred. When we tested you, you only could differentiate gray values of zero to fifty. Obviously the machine is more sensitive.”

No, you are not reading the script of the Hollywood movie based on this thesis. The conversations are taken from *Brain*, written by Robin Cook, the godfather of the medical thriller, in 1979. It’s probably the only bestseller about computers reading radiographs ever written.

This thesis is another book about computers reading radiographs.

But I’m afraid it will never become a bestseller because the plot is not so spectacular. Unlike William Michaels, I will not turn out to be a modern Frankenstein who kills young beautiful women, puts their brains in cylinders with cerebrospinal fluid and somehow hooks up a keyboard and monitor to the floating gray matter. Instead I used meagre Intel Pentium processors and this comes at a price: the performance of my programs is nowhere near that of *RadRead*. Nevermind the blatant nonsense it contains, *Brain* has a point here: if we unravel the mechanisms that underlie visual perception in living systems, we may be able to program computers to display similarly impressive performance.

In fact, this idea attracted me to this field, to the research group I have been working in, and to the topic of my PhD project. I’ve learned many things about image processing, with techniques that are firmly based in physics and mathematics and that are already powerful enough to be used successfully in real applications. On the other hand, I have also arrived at the opinion that many expectations and predictions about the capabilities of computer vision and artificial intelligence systems, that must have inspired Robin Cook to write his *Brain*, are totally unrealistic. To be a little bit more precise, articles from the 1970s on computer analysis of radiographs talked about *automating the chest exam*. This did not happen in the past thirty years, and I do not believe it will happen in the next thirty years.

I am aware that I just tried to predict the future. And predicting the future is tricky business. Consider the following prediction from the famous English logician Alan Turing. He put forward a practical definition of a thinking machine. An interrogator is asking questions to someone in a different room. The conversation is only displayed on a computer monitor. If the interrogator cannot deduce whether the answers are given by a person or a computer imitating a person, the computer can be considered a thinking machine. In an article that appeared in 1950 [238] Turing states:

“I believe that in about fifty years’ time it will be possible to programme computers, with a storage capacity of about 10^9 , to make them play the imitation game so well that an average interrogator will not have more than 70 per cent chance of making the right identification after five minutes of questioning.”

Turing was wrong, but the amazing thing about this quotation is that the casually made second prediction, about the existence of computers with a storage capacity in the order of 10^9 around the year 2000, is correct! This is usually referred to as Moore’s Law, formulated thirty years ago by Intels co-founder Gordon Moore. He stated that computer processing power could be doubled every 18 to 24 months. The chip industry has been able to live up to this incredible rate of progress ever since and is expected to continue to do so for at least the next decade or so. I believe that Moore’s Law is the prime explanation for the progress made in computer vision and, more specifically, computer-aided diagnosis. Obviously, the efforts of image processing researchers were essential too. There is now a commercial product available that assists radiologists in reading mammograms, the **ImageChecker** from R2 Technology, Los Altos, CA. I believe such products will become available for chest radiography as well, and that this will increase the effectiveness and efficiency of medical care. And I hope that the research described in this thesis will contribute to this.

Chapter 1

Introduction and Outline

FIGHT THE OLD ENEMY TB says the stamp on the next page, published to commemorate the discovery of the bacillus that causes tuberculosis (TB). It shows a chest radiograph of someone infected with TB. This thesis is about the development of tools for the computer analysis of chest radiographs, notably those obtained in mass chest screening against TB. Currently digital chest units are available for mass screening, such as the Oldelft Digidelca manufactured by the Dutch company Nucletron, a partner in this research project. Such units no longer require the use of films: all images are directly available in digital form. Computers could analyze these images without the need for digitization and aid the radiologist in reading them. In this thesis, several new methods are described for segmenting chest radiographs and for texture analysis that can be applied to detect abnormal signs in these images. The application of these methods is not limited to mass screening; they can be applied to clinical chest radiographs as well.

This chapter provides some background information on tuberculosis and chest radiographs and gives an outline of the thesis.

Tuberculosis

The World Health Organization (WHO) has declared tuberculosis a global emergency. TB is estimated to kill three million people each year. Yet most people infected with the germ, *Mycobacterium tuberculosis*, do not suffer from TB. Instead the germ becomes trapped in body tissues, sealed up in a calcified nodule, or *tubercle*. On a radiograph such a nodule can produce a tiny dense shadow at the edge of the lung field, often accompanied by a shadow representing lymph nodes in the center of the field. These two shadows represent the primary complex of tuberculosis and such an infection is common in children, who often do not suffer from illness and may develop immunity to TB. In adults, TB usually takes a different form. The bacillus spreads widely in the lungs and forms large cheesy (caseous) masses that break down the respiratory tissues and develop cavities in the lungs. Once the disease process eats its way into the bronchi, the bacilli can be coughed or breathed out, making TB highly



Figure 1.1: Three stamps that show the importance of chest radiographs and mass screening in campaigns against tuberculosis. These images were obtained from the web-site *A Century of Radiology* at <http://www.xray.hmc.psu.edu/rci> which includes a philatelic history of radiology.

contagious. A blood vessel can be eroded and then the patient coughs up blood. This spreading type of the disease was popularly known as *consumption*, a name that vividly describes TB's destructive progress. If the patient is treated with drugs, or builds up resistance, the bacilli may be sealed up in the lungs, which by then contain much scar tissue. Consequently, the disease can, in its various stages, give rise to a large variety of signs on an X-ray.

Accounts of tuberculosis can be found as far back as the writings of the ancient Egyptians. TB was the leading cause of death for all age groups in the Western world from that period until the early 20th century. Finding an effective cure for tuberculosis represents one of the great advances in 20th-century medicine. Fifty years after Robert Koch, in 1882, discovered the tubercle bacillus, several antimicrobial drugs were discovered that can cure the disease.

The WHO [193, 171] now advocates DOTS therapy, which stands for directly observed treatment in which the patient takes a strong cocktail of drugs, under direct observation of medical personnel, which can eradicate the bacilli in a short period of time. This strategy is adopted to ensure that patients finish their treatment. Otherwise, bacilli resistant to several drugs multiply and the patient becomes sick again. These multi-drug-resistant (MDR) strains of bacilli cause an acute form of the disease that is extremely difficult to cure and in most cases proves fatal. Outbreaks of MDR TB in the United States and Europe have focused attention on TB as a worldwide threat instead of a problem of developing countries. Another major cause of concern is the deadly combination of AIDS and TB. If people infected with HIV subsequently become infected with TB – and TB is probably the most important single cause of morbidity among this group – they have an extremely high risk of developing active TB within a short period of time. Therefore HIV infection can *telescope* a TB epidemic, shortening the time to generate the epidemic from years to

months.

Chest radiographs have always played an important role in differential diagnosis and determining the extent of tuberculosis, although they are not the only way to diagnose the disease. A problem is that no radiological sign is unique to TB and that an abnormality does not prove the presence of active disease [85].

Chest radiography

X-rays were discovered by Conrad Wilhelm Röntgen in 1895 [208], while he was experimenting with electric current flowing in a cathode-ray tube. He noticed the cathode emitted some unknown radiation that could be made visible with photographic paper and that passed through many materials. He did not realize the radiation was a form of light and therefore called them X-rays. His discovery heralded the age of modern physics and revolutionized diagnostic medicine. Getting a chest radiograph has become and still is the most common radiological exam [29].

A tutorial on chest radiography is beyond the scope of this chapter, but without some minimal knowledge of chest images, some parts of this thesis will be hard to appreciate. Therefore two figures are included here. Figure 1.2 shows a normal chest radiograph with several normal structures labeled. Figure 1.3 shows a number of close up views of parts of abnormal chest images. The displayed findings are by no means complete or statistically representative, but serve only as an illustration of the variety of abnormalities that can be encountered on a chest X-ray.

It is not unlikely that a reader of this thesis, especially if he or she is not a radiologist, cannot see a detail or an abnormality in the image that is referred to in the text. This can happen especially for the smaller images in the printed version of this thesis. The quality of the images in the electronic version, when viewed on a (standard) computer monitor, is much better. And even that quality is clearly inferior to the average film hanging on an average light box. Moreover, it is good to keep in mind that even the judgements of radiologists on abnormalities in chest radiographs turn out to be all but consistent in practice [273,77]. Yerushalmy describes the results of several early studies:

After two years of work the CRN group reluctantly had to admit that “the present writers appear to have failed in this project of finding a reliable classification of the röntgenographic appearance or quality of a tuberculous pulmonary lesion. It is believed that reasonable degrees of tenacity and resourcefulness were used in the attempt. It was disappointing to find that many conferences and much practice, together and apart, failed to increase the reliability and agreement to a useful degree.” [...] The challenge was taken up enthusiastically by the GMZ group. [...] Unfortunately they were no more successful than the original group. They stated, “At least two members of GMZ had felt that *they* knew whether a lesion was homogeneous or not, whether it was fuzzy or sharply defined, and what its shape was . . . but experience awakened them.”

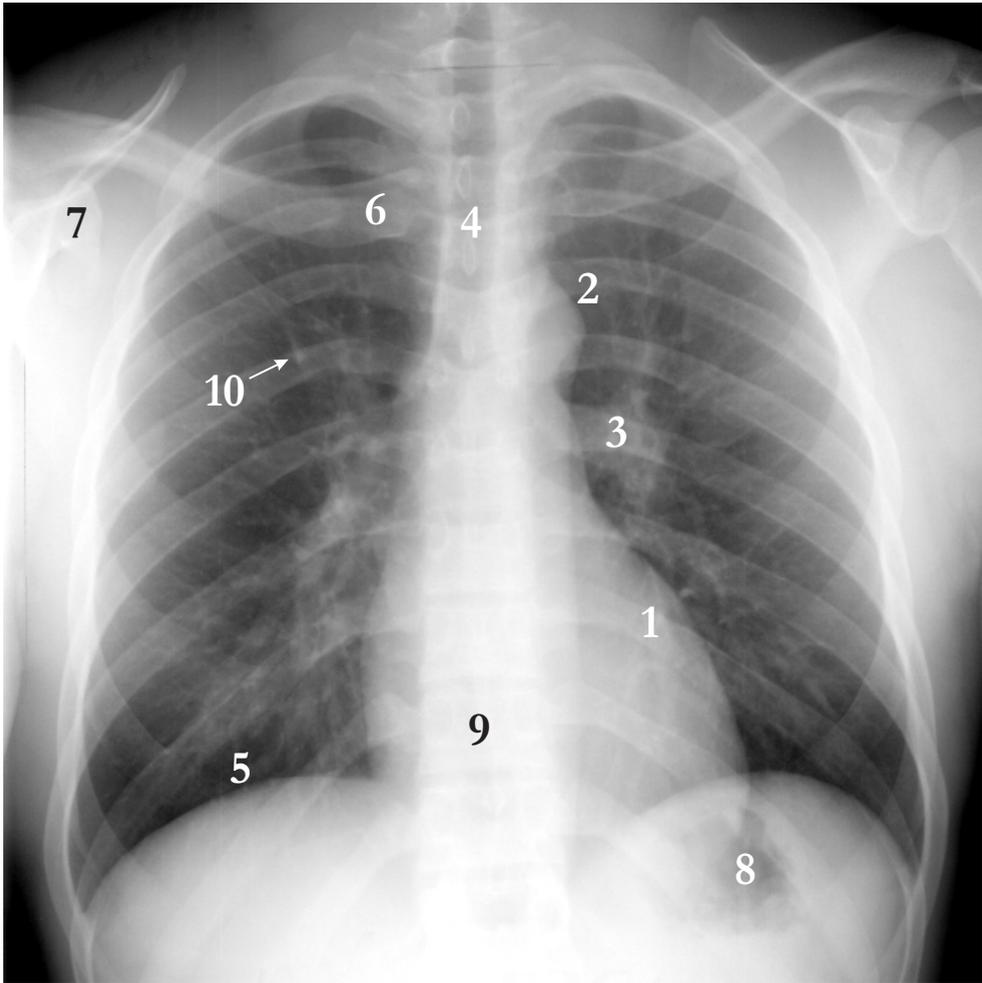


Figure 1.2: A normal PA chest radiograph (PA stands for posterior-anterior which means that the radiation passes through the patient from back to front). The patient always faces the observer: the left side of the image shows the right lung. The lungs are radiolucent baskets of air, therefore they show up black in the image (by convention brightness indicates absorbed radiation). Within the lung fields, only bony structures and blood vessels are visible. The posterior ribs (in the back of the patient) are visible more clearly; you can follow them “turn” into the anterior ribs. 1) The heart. 2) The aortic arch where the aorta bends. 3) The (in this case left) hilum, where the arteries and veins enter the lung. 4) A darker vertical stripe indicates the trachea. 5) Below the lung fields the diaphragm starts. 6) Clavicle. 7) Shoulderblade. 8) Usually stomach gasses can be seen in the left diaphragm. 9) If there is enough contrast in the mediastinum (the area projected between the lung fields) the spine may be visible. 10) A round dot like this is the shadow of a vessel that runs in the same direction as the X-rays. So this is a normal finding. But distinguishing such end-on vessels from abnormal signs, such as nodules, is sometimes very difficult.

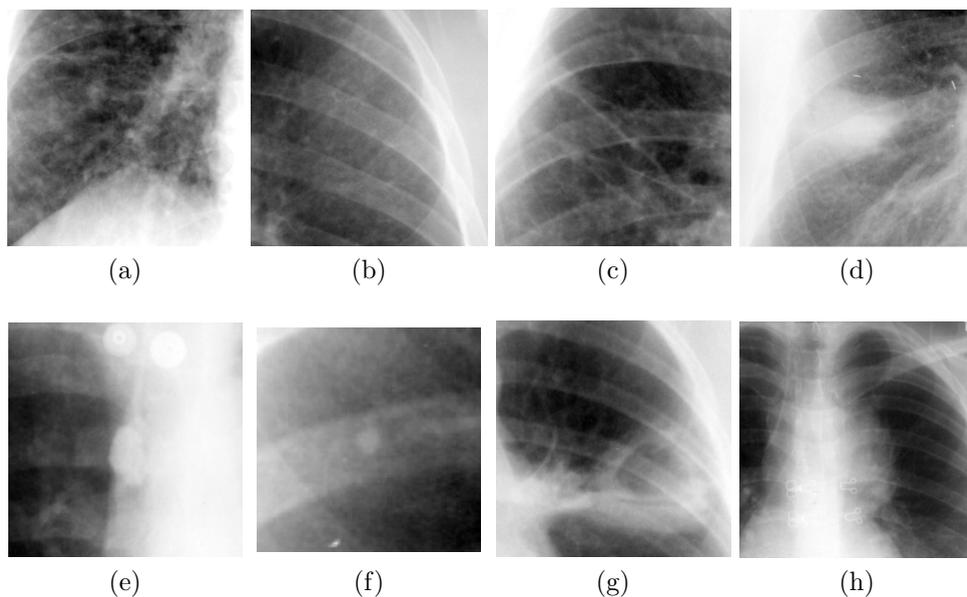


Figure 1.3: (a) Obvious diffuse abnormalities in the lower right lung. (b) More subtle diffuse abnormalities. (c) Linear scarring, there are abnormal lines that run from top-left to lower right. (d) The denser area in the center is an obvious infiltrate. (e) The round dense object in the middle is a calcified lymph node. Note the two buttons in the top, such artifacts can mimic abnormal structures. (f) A small calcification located behind a rib. (g) A cyst with a cavity. (h) The hilum of the left lung is abnormal; the lymph nodes have swollen. This indicates primary TB.

So we conclude this brief introductory section by stating that interpreting chest radiographs is a hard problem, for both computers and radiologists.

Outline of this thesis

Our general strategy to analyze chest images can be briefly stated as follows. The image is automatically segmented and divided into regions. The texture in these regions is quantified and compared with texture in corresponding regions of images in a database. Based on this comparison the abnormality of each region can be estimated. These estimates can be aggregated into a probability that the image is abnormal. Segmentation and texture analysis are therefore the two main threads in this thesis. Some chapters are methodological and approach segmentation and texture analysis from a general point of view. Others are directly about the application of segmentation and texture analysis techniques to chest radiographs.

Chapter 2 reviews the literature on computer analysis of chest radiographs and proposes directions for future research. Chapter 3 describes several automatic methods to segment the lung fields in chest radiographs and compares their performance

in experiments on a large number of chest radiographs. Chapter 4 introduces locally orderless images, a framework to describe images in terms of local histograms, and demonstrates how this framework can be the basis for several common image processing tasks. Based on a generalization of locally orderless images, a new method for texture feature extraction is introduced in Chapter 5, and tested in experiments on a collection of texture images. Chapter 6 is about an extension of active shape models, an established general segmentation method. We show how one can use local texture features, based on the framework laid out in the preceding chapters, to find optimal descriptors that can steer the segmentation and outperform the original active shape models in segmentation tasks in chest radiographs and slices from MRI brain data. In Chapter 7 a statistical model of the complete rib cage is constructed and a method to fit this model automatically to chest radiographs is described and tested. The basis of this rib cage model is principal component analysis, on which active shape models are also based. Finally, in Chapter 8, the segmentation and texture analysis techniques are combined into a single method that automatically detects textural abnormalities in chest radiographs and estimates the probability that an image is abnormal. The method is tested on a database of clinical chest films with interstitial disease and on a database from a TB screening program. The last chapter summarizes the thesis and discusses the overall results.

Chapter 2

Computer Analysis of Chest Radiographs — A Review

The literature on computer analysis of chest radiographs is categorized and briefly reviewed. Some directions for future research are given.

THE DISCOVERY OF X-RAYS by Wilhelm Conrad Röntgen, in 1895 [208], has revolutionized the field of diagnostic medicine. The chest radiograph is the generic example of “an X-ray image”. And rightly so because even today, with a wide variety of (three-dimensional) imaging techniques available, the traditional two-dimensional chest radiograph remains ubiquitous in clinical practice, making up about a third of all exams in a typical radiology department [29].

Soon after the invention of the modern digital computer at the end of the 1940s, research began on having computers perform tasks that had previously been performed only by human intelligence. Computer-aided diagnosis – which may be defined as a diagnosis made by a radiologist who takes into account the result of a computer analysis – is one domain of this huge field of artificial intelligence. The first articles about computer analysis of radiographic images appeared in the 1960s [16, 182]. Articles describing techniques specifically designed for computerized detection of abnormalities in chest radiographs began to appear in the 1970s. These were among the first papers in the field of medical image analysis, as is noted in a review by Duncan and Ayache [64].

These early studies displayed a considerable optimism regarding the capabilities of computers to generate complete diagnoses. They are summarized in one review [44] as attempts to *fully automate the chest exam*. Over the decades this expectation has subsided, as seems to have been the case with the early enthusiasm regarding the capabilities of artificial intelligence systems in general. Currently the general agreement is that the focus should be on making useful computer-generated information available to physicians as they make decisions instead of trying to make a computer act like a diagnostician [66]; from the abbreviations FACD (fully automatic computer diagnostic

systems), ICD (interactive computer diagnostic systems), and CAD (computer-aided diagnosis), the latter appears to be the only surviving acronym. But a discussion on whether computers should independently make diagnosis or only aid the radiologist is not the main topic of this chapter; we merely aim to categorize and briefly review the literature on computer analysis of chest radiographs with emphasis on describing the techniques that have been employed and the tasks these techniques are supposed to solve. Care has been taken to make this review as complete as possible. The focus is on papers in international journals. Chapters from conference proceedings have been added if they contained work that has not been described in journal articles.

Literature on acquisition, about digital chest units versus analog film systems (e.g. [40,214,74,45]), or on dual energy systems in chest radiography (e.g. [107,67,90]) is beyond the scope of this review as are studies that use psychophysics, e.g. to determine the optimum tube voltage [10] or to aid the detection of abnormalities by measuring visual dwell [106,144]. Similarly we shall not go into detail about studies that quantify the performance of radiologists, usually through observer studies, when image quality parameters of the chest radiographs are varied (e.g. [170,169,166,254,46,50,210]). Furthermore, research on classifiers that estimate probabilities that patients exhibit a certain disease, given a number of features from clinical information and/or output of computer algorithms is largely ignored here since this work belongs to the field of pattern classification instead of medical image processing. One of the earliest examples in this area is the work of Lodwick *et al.* in 1963 [165] who estimated one year survival statistics for lung cancer patients based on characteristics from chest radiographs. Such studies have continued to appear, e.g. [231,9]. In a recent paper Ashizawa *et al.* [11,12] trained a neural network to estimate the likelihood of 11 diseases based on 10 clinical parameters and 16 radiological findings.

When trying to subdivide the literature, we discern three main areas: general pre-processing techniques, algorithms for segmentation of anatomical structures and, finally, analysis aimed at solving a particular task or application, usually an attempt to detect a specific kind of abnormality. The following subdivision is adopted and followed in the sections of this chapter:

- General processing
 - Enhancement
 - Subtraction techniques
- Segmentation
 - Lung fields
 - Rib cage
 - Other structures
- Analysis
 - Size measurements
 - Lung nodule detection
 - Texture analysis
 - Other applications

Conners *et al.* [44] wrote an interesting review on computer analysis of radiographic images in general in 1982, with the emphasis on chest radiography. Overviews of research at the Kurt Rossmann Laboratories at the University of Chicago, one of the most active groups in computer-aided diagnosis, are given in [1,60,59,61,57,62,58].

Almost all work is applied to frontal (posterior-anterior or PA) chest radiographs. A few studies are aimed at lateral radiographs, and those cases are specifically mentioned.

The enormous amount of information that radiologist seemingly casually extract from radiographs may give a computer scientist, working on algorithms for the analysis of chest radiographs, a sense of despair. But this should not stop him or her from obtaining at least some knowledge about the anatomy of the chest, its appearance on projection radiographs and the nature of various abnormal findings. The author enjoyed the introductory textbook by Squire [225] and the part on differential diagnosis of chest radiographs in Lange's standard textbook [149]. A detailed pictorial description of normal and abnormal findings in chest radiographs is given in the chapter on chest imaging in [263]. A nice overview of interstitial disease can be found in [262].

Several tables that summarize papers on specific topics are included. Studies are listed in chronological order in these tables. In most cases we included information about the evaluation of the presented techniques, such as the size of the databases that are used. We do not include reported figures on classification performance or segmentation accuracy, because these results are usually very dependent on the subtlety of the abnormalities in the processed images which makes it impossible to compare the merits of various methods on the basis of such figures alone.

General processing

Enhancement

Chest radiographs inherently display a wide dynamic range of X-ray intensities. In conventional, unprocessed images it is often hard to “see through” the mediastinum and contrast in the lung fields is limited. A classical solution to this kind of problem in image processing is the use of (local) histogram equalization techniques. A related technique is enhancement of high frequency details (sharpening). A variety of pre-processing techniques for chest radiographs based on local equalization, sharpening and combinations and modifications of these techniques have been proposed [98, 41, 176, 180, 218, 109, 21, 230], sometimes not only through software but also by using hardware [23, 78]. Nowadays almost all vendors sell digital chest units with a larger dynamic range than conventional units and most vendors automatically preprocess the images with proprietary algorithms.

Subtraction techniques

Subtraction techniques attempt to remove normal structures in chest radiographs, so that abnormalities stand out more clearly, either for the radiologist to see or for computer analysis to detect.

The first approach is temporal subtraction, proposed by Kano *et al.* [124]. A previous radiograph of the same patient is registered with the current image. An elastic matching technique is employed in which the displacement of small ROIs is computed based on cross-correlation and a smooth deformation field is obtained by fitting a high order polynomial function to the displacement vectors. The registered image is subtracted and if the registration is successful, areas with interval change appear as either dark or bright on a gray background. The original technique has been improved and evaluated using subjective ratings by radiologists of the quality of the subtraction image [110, 113] and the results of the method have been compared with manual registration [131]. An observer study [56] with a small number of selected cases showed an increase in detection accuracy of interval change when both the normal image and the subtraction image were presented to the radiologist. With the advent of digital archives, temporal subtraction techniques may be applied on a routine basis since monitoring interval change is one of the main reasons for making chest radiographs. It is common to obtain a chest radiograph of patients on an intensive care unit every day.

If a previous radiograph is not available, a subtraction can be made by mirroring the left/right lung field, performing elastic registration on the right/left lung field and subtracting. This technique, coined *contralateral subtraction*, employs the symmetry of the rib cage. Li *et al.* [155, 154] use a scheme similar to Kano *et al.* [124], based on cross-correlation and evaluate the result with subjective ratings of the quality of the subtraction image by radiologists. Yoshida [274, 275] minimizes the sum of squared differences, using a mapping that is ensured to be smooth through a regularization term and the optimization is done in the wavelet domain. He uses the technique to eliminate false positives from a number of candidate lung nodules.

It may also be possible to eliminate or suppress normal structures in radiographs by subtracting a model that is fitted to the input image. The technique to remove ribs by Vogelsang *et al.* [253, 252] is an example of this approach. Ribs borders are detected and a simple physical model of the density of the rib cage is subtracted from the input image. The technique is not evaluated.

Segmentation

Lung fields

Automatic segmentation of the lung fields is virtually mandatory before computer analysis of chest radiographs can take place. Several studies deal with this problem exclusively. Others deal with parts of the problem, such as detection of the outer ribcage [269], the diaphragm [270], or the costophrenic angle (where the diaphragm and the rib cage meet) [7]. A few papers are focused on lateral chest radiographs [4, 32].

Table 2.1 gives an overview of papers on lung segmentation. Broadly speaking, there are two approaches: rule-based reasoning and pixel classification. A rule-based scheme is a sequence of steps, tests and rules. Most algorithms for the segmentation of lung fields fall in this category [198, 269, 270, 65, 4, 6, 7, 32, 82, 84]. Techniques employed are (local) thresholding, region growing, edge detection, ridge detection, morphologi-

	RB	PC	PA / lateral	evaluation	remarks
Toriwaki [233, 232]	×		PA	none	describes a complete analysis system
Harlow [102]	×		PA	none	
Chien [37, 38]	×		PA	none	only right lung, result used to detect abnormalities
Hasegawa [103]		×	PA	none	
Pietka [198]	×		PA	none	
McNitt-Gray [179, 178]		×	PA	16 Q	uses 5 anatomical classes
Duryea [65]	×		PA	802 Q	
Xu [269]	×		PA	1000 R	outer rib cage only
Xu [270]	×		PA	300 R	diaphragm edges only
Armato [7]	×		PA	600 R	costophrenic angles only
Armato [4]	×		lateral	200 Q	
Armato [6]	×		PA	600 R	
Carrascal [32]	×		both	65 RQ	
Vittitoe [250]		×	PA	99 Q	
Tsujii [235]		×	PA	71 Q	
Wilson, Brown [265, 28]	×		PA	none	describes a complete analysis system
Vittitoe [251]		×	PA	115 Q	uses 6 anatomical classes
Van Ginneken [82, 84] (Chapter 3)	×	×	PA	115 Q	
Van Ginneken (Chapter 6)			PA	115 Q	general knowledge-based method

Table 2.1: Some attributes of papers on lung field segmentation. The first column lists the first author and reference(s). Methods are divided in rule-based (RB) and/or pixel classification (PC) schemes, see the text for details. The column evaluation gives the number of images used in evaluation only (images used in a test set are not included). R means evaluation through subjective rating by a radiologist of the result, usually in classes from 1 (poor) to 5 (excellent). Q stands for quantitative evaluation, in which the result is compared on a pixel-by-pixel basis with a manual segmentation.

cal operations, fitting of geometrical models or functions, dynamic programming. On the other hand, several attempts have been made to classify each pixel in the image into an anatomical class (usually lung or background, but in some cases more classes such as heart, mediastinum, and diaphragm are used [178, 251]). Classifiers are various

types of neural networks, or Markov random field modeling, trained with a variety of (local) features including intensity, location, texture measures [179, 178, 250, 251, 235]. Van Ginneken and ter Haar Romeny [82, 84, Chapter 3 of this thesis] combine both approaches in a hybrid scheme.

It is also possible to use general knowledge-based segmentation methods, such as active shape models [49] for the segmentation of lung fields, as is shown in Chapter 6 and 8 of this thesis.

In studies done in the 1970s often a rough or partial outline of the lung fields was detected, with rule-based schemes usually based on analysis of profiles [143, 227, 194, 97, 115]. In this chapter, the term “profiles” refers to an average of consecutive 1-D lines of pixels, usually running horizontally or vertically. Evaluations of the effectiveness of these methods are not made, and none of these papers focuses on lung field segmentation specifically but use the segmentation for further processing or to estimate the heart size or lung capacity. The methods presented in these papers undoubtedly inspired later researchers.

On the whole, the problem of segmenting lung fields has attracted considerable attention. Several authors report good or excellent results that approach the inter-observer variability. The task might therefore even be considered largely solved, although no attempts have been made to test methods with very large databases to verify if schemes are also able to produce reasonable results for the, say, 1% of most difficult cases. Likewise, a study in which the performance of various methods on a large common database of radiographs (of varying quality and obtained with different settings) is compared has not been made and would be an interesting endeavor. In Chapter 3 of this thesis, some quantitative results of several studies, although obtained from different datasets, are reported.

Rib cage

There are several reasons why the automatic delineation of rib borders can be useful for computer analysis. First of all, the ribs provide a frame of reference and the locations of abnormalities are often indicated by radiologists in terms of numbered (inter)costal spaces. Second, rib border segmentation may be used to detect rib abnormalities such as fractured or missing ribs. Third, once locations of ribs are known, regions of interest between (in the intercostal space) or on (in the costal space) the ribs can be defined for further analysis. Fourth, knowledge about the location of rib borders may be used to eliminate false positives in the detection of abnormalities such as nodules. Rib crossings (of the posterior and anterior parts of the ribs) frequently turn up as false positives in nodule detection schemes.

The approaches to rib (cage) segmentation are summarized in Table 2.2. Only papers with a clear description of the algorithm have been included. Note that several studies also detect (parts of) the contours of the lung fields [55, 201, 278] but do not evaluate these methods. The classical approach, which is followed by most researchers, is the following. A geometrical model of rib borders is chosen, e.g. parabolas or ellipses or some combination. Edge detection provides candidate locations for these rib borders. Either the candidates are rejected or grouped, and the model is fitted to candidates, or each candidate votes for combinations of model parameters (the latter

	geometric model	detection method	P/A	evaluation
Toriwaki [233, 232]	parabolas	edge detection, parabolas fitted through candidates, candidates are merged	PA	15A, 40A
Wechsler [256, 258, 257]	parabolas, ellipses, 4th order polynomials	oriented edge detection, modified Hough transform and rules to link segments	PA	5A
Ballard [14]	–	models relations between ribs to classify rib border candidates	P	4A
De Souza [55]	a combination of two parabolas	vertical edge detection, grouping, rejecting, refining candidates and fitting	PA	8A
Powell [201]	detects small square ROIs in inter-rib spaces	fits a periodic function with varying frequency for costal and intercostal space to pixel data	–	66Q
Sanada [211]	ellipses with height given by function from [201]	vertical edge detection, grouping candidates and fitting geometric model	P	50R
Yue [278]	parabolas, snakes	edge detection, modified Hough transform, segment selection rules, followed by snake fitting	P	10Q
Vogelsang [253, 252]	parabolas, snakes, active shape models	edge detection, template matching, modified Hough transform, rules to select among candidates	P	154A
Van Ginneken [83] (Chapter 7)	parabolas, statistical rib cage model from PCA	optimization of model parameters with cost function based on oriented edge detection	P	35Q

Table 2.2: Summary of papers on rib detection in frontal chest radiographs. The first column lists the first author and reference(s). The column P/A indicates if posterior (P) or anterior (A) ribs are detected, or both or none. The right column lists on how many images the method is evaluated and if the evaluation is a subjective judgement from the author (A), based on ratings by radiologists (R) or quantitative (Q).

approach is referred to as a modified Hough transform). Post-processing steps remove some borders, or infer new rib borders that had not been detected. In some cases a fine-tuning stage is added in which the geometric model may be abandoned, such as the snakes fitted to rib borders in [278].

A difficulty in the classical approach is that the relations between ribs are not implicitly taken into account during the fitting procedure. Ribs may be missed or detected twice and the fact that consecutive ribs and left-right ribs have similar shape is often ignored. Early attempts to model the relation between ribs, as in [14] do not appear very powerful or use very simple ad hoc rules [278]. This was already recognized by Wechsler [257] in 1977: “We believe that at this stage a significant improvement in rib boundary detection could be achieved only by a knowledge-based system which would incorporate a model of the rib cage”. The rib cage model in [83] (Chapter 7 in this thesis) is an attempt to model the shape of the complete rib cage and fit it directly to the image instead of rib-by-rib sequentially.

The number of images used in evaluation is in general considerably smaller than in the case of lung field segmentation (compare Table 2.2 with Table 2.1) and in many cases the evaluation is confined to qualitative judgements from the authors themselves (these studies are indicated with A in the right column of Table 2.2). Since the full rib cage has a much more complex shape than the outline of the lung fields, partial failures are more likely. And they occur frequently, as is evident from the discussions in various articles. Furthermore, the detection of anterior (or ventral, as opposed to the posterior or dorsal) parts of the ribs has been addressed only in studies that omit proper evaluation. Therefore we conclude that the problem of automatic segmentation of the rib cage in chest radiographs is still far from solved.

Other structures

In this section we mention work on segmentation that does not fit into the lung field or rib cage segmentation proper.

Toriwaki *et al.* describe a complete system for the analysis of chest radiographs [233,232] that includes segmentation of lung fields, rib cage, heart, clavicles and blood vessels followed by automatic detection of abnormalities. Unfortunately, only a rough description of results on 15 [233] and 40 images [232] taken from a mass screening for lung tuberculosis in Japan is given and there is no mention about the nature and subtlety of the abnormalities in this database.

A method to select ROIs within the lung fields that do not contain rib borders is presented by Chen *et al.* [36]. The lateral lung fields are subdivided into regions which are partly eliminated if they contain edges of certain strength and orientation.

Segmentation of (parts of) the heart boundary has been studied by several researchers [182,96,143,217,189,188,252], usually with the aim of detecting cardiomegaly (enlarged heart size). Note that a correct segmentation of the lung fields is sufficient to compute the cardiothoracic ratio indicative of cardiomegaly, since parts of the boundaries on the lung fields coincide with the heart contour. In the section on size measurements these studies will be discussed in some more detail.

Armato *et al.* [5] developed an algorithm to detect abnormal asymmetry in chest radiographs using a rule-based scheme for detecting lung contours and comparing the projected lung areas. The method was evaluated on 70 radiographs.

Analysis

What kind of abnormalities does the radiologist who reads chest radiographs encounter in his/her day-to-day practice? It seems that the answer to this question should set the agenda for computer scientists who want to – as it was put in the introduction to this review – “fully automate the chest exam”. MacMahon [168] has attempted to answer this question by counting the abnormalities encountered in 1085 abnormal chest radiographs, collected from consecutive cases, and dividing them in 30 categories. The main results are summarized in Table 2.3. The statistics are surprising once they are compared with the kind of abnormalities that are commonly investigated in the literature on computer analysis of chest images. Lung nodules are a rare abnormality but receive much attention in CAD literature and observer studies tend to use their detectability as a criterion for diagnostic accuracy. Pulmonary infiltrates could be detected in principle by the texture analysis schemes to be discussed later and are the most commonly occurring abnormality. Catheters are next, and although catheters usually stand out clearly in the image, locating their tip was the task that was most frequently limited by image quality. One may argue that the presence of a catheter, or drainage tube or pacemaker is not an abnormality. However, the presence of such objects can have a large effect on computer analysis and it is important to realize how common such findings are in abnormal radiographs. The same is true for clothing artifacts.

The following applications of computer analysis of chest radiographs are reported in the literature:

- Lung nodule detection.
- Detection of cardiomegaly or estimation of the cardiothoracic ratio.
- Estimation of total lung volume.
- Detection of pneumothorax.
- Estimation of the severity of pneumoconiosis (coal worker’s disease).
- Detection of interstitial disease.
- Detection of abnormalities encountered in mass screening for tuberculosis.

We divide these applications into four groups. Size measurements are applications based on results that can be computed directly from a segmentation. The detection of nodules is considered a separate category. Several other applications are grouped under “texture analysis” and the remaining tasks are listed under “other applications”.

Size measurements

In a few applications, the results can be obtained directly from segmentations. An example is the estimation of enlarged heart size (cardiomegaly) usually performed by measuring the cardiothoracic ratio (CTR), the maximal horizontal diameter of the heart divided by the maximal horizontal diameter of the thorax. This is the

Finding	% of all abnormalities
Pulmonary infiltrates	55%
Intravenous catheters	33%
Heart size/contour	27%
Endotracheal/tracheostomy tubes	22%
Pleural effusions	12%
Linear atelectasis/scar	10%
Drainage catheters and tubes	9%
Pulmonary vascularity	9%
Pleural scarring	8%
Rib lesions	7%
Mediastinal masses	6%
Diaphragm	5%
Calcified granulomas	5%
Pneumothorax	5%
Lung nodules	5%
Extrathoracic abnormalities	4%
Lung masses	2%
Calcified nodules	1%
Mediastinal shift/contour	1%
Cardiac pacemakers	1%

Table 2.3: The frequency with which abnormalities are encountered in abnormal chest radiographs encountered in clinical practice. Data was obtained from 1085 standard radiographs of adults. The information in this table is taken from the study “The nature and subtlety of abnormal findings in chest radiographs” by H. MacMahon, K.J.M. Liu, S.M. Montner, and K. Doi [168].

subject of probably the earliest studies of computer analysis of chest radiographs by Becker, Meyers, and co-workers [16, 182]. They detected points on the rib cage boundary and heart boundary using horizontal profiles and estimated the CTR of 37 radiographs. In work by Hall [96], Kruger [143] and their co-workers, several parts of the lung and heart boundaries were segmented and various ratios were computed. Entering these ratios as features in linear and quadratic discriminant functions, they attempted to classify images as normal, showing rheumatic heart disease or containing other abnormalities. The method was evaluated on 320 films, a surprisingly large number for studies from this period. Using a similar approach, Sezaki and Ukena [217] computed the CTR by proposing a scheme that detects the vertical boundary of the rib cage and the heart through analysis of horizontal profiles and the application of a few rules to correct failures. Roellinger *et al.* [206] used the method from [143] to detect points on the heart boundary and fitted a Fourier shape to them. The coefficients of the Fourier shape were used to classify the heart shape as normal or abnormal. The classification was correct in 72% of all cases and a database of 481 images with 209 normals was used. A comparable method was described by Nakamori *et al.* [189]. A Fourier shape was fitted to points on the heart boundary found by edge

detection and the results were used to detect cardiomegaly in [188] and evaluated on a database of 400 radiographs with 91 abnormalities. All in all, this problem has received considerable attention in the 1970s and the results are promising. It is likely that modern segmentation methods will outperform the schemes used in most studies cited here since these are based on simple rules to detect only a few landmark points in the image. Clinical applications of automatic detection of heart shape abnormalities seem feasible, but the problem has not attracted much attention the last ten years.

A related example of size measurements is the determination of total lung capacity (TLC). In this case boundary points on both a PA and a lateral radiograph must be detected. The volume is estimated using formulas proposed by clinicians that assume simple shapes for the lungs and have been shown to correlate well with the true TLC. Paul *et al.* [194] described a method to automatically determine the TLC using profile analysis similar to Kruger *et al.* [143]. It was tested on 15 radiographs. Carrascal *et al.* [32] used their rule-based segmentation of PA and lateral lung fields to estimate the TLC for 65 radiographs.

Nodule detection

Lung cancer is a major cause of mortality in the Western world. Early detection of lung tumors (visible on the chest film as nodules) may greatly increase the patient's chance of survival. Strikingly, retrospective studies showed that 90% of peripheral lung cancers were visible on earlier radiographs [186]. Therefore this problem has received a lot of attention and is the most studied problem in computer analysis of chest radiographs.

Nodules show up as bright (although the contrast may be very small) round objects within the lung fields. The difficulty, especially for CAD schemes, is to distinguish true nodules from (overlapping) shadows from vessels and ribs.

Almost all researchers use a two step approach for nodule detection. In the first stage candidate nodules are detected. The second stage consists of eliminating as many false positive candidates as possible, without sacrificing too many true positives. This approach is reminiscent of the two tasks of a radiologist in the diagnostic process: identifying the location of abnormalities and categorizing them as benign or representing a specific type of disease.

Table 2.4 lists papers that describe algorithms for the detection of candidate nodules. Tables 2.5 and 2.5 list work on eliminating false positive candidates. Note that several studies discuss both steps but focus on one of them in particular. These papers show up in only one table.

Several schemes start with producing an image in which nodules are enhanced (by filtering with a nodule-like filter, essentially template matching) and/or background structures suppressed (by some sort of blurring) [79, 163], or other pre-processing steps are applied, such as unsharp masking or similar operations [147]. Subsequently candidates are detected using Hough transform techniques in which edge pixels vote for circles that could cause these edges [13, 15, 147, 148, 173], or template matching is used or the nodule enhanced image is simply thresholded. Some studies apply a background trend correction on ROIs with nodule candidates [271, 174], which is similar to high pass filtering of the input image as a preprocessing step.

study	method	evaluation
Sklansky [221], Ballard [13, 15]	Generalized Hough transform: edge detection, edges vote for centers of circles of different radii, using gradient magnitude. Non-maxima suppression.	6 images
Sklansky [222]	Spoke filter, comparable to [13, 15].	19 images, 16 abnormalities
Lampeter [148]	A blurred version of the image is subtracted. The result is histogram equalized. Candidates are detected with a Hough transform, similar to [13].	37 images
Giger [79]	Subtraction image is produced by filtering with a spherical kernel and subtracting a median filtered version. Nodules candidates are found by thresholding.	real and simulated nodules
Yoshimura [277]	Subtraction image is produced by a morphological open operation and subtracting a median filtered version. Compared and combined with [79].	60 images, 30 abnormalities
Suzuki [229, 228]	A non-linear filter with three circular bands subdivided in segments produces output if the maximum segment values of each band decrease.	192 nodules
Lo [163]	A subtraction image is constructed similar to [79]. On the subtraction image, template matching with cross-correlation and sphere profiles is performed, followed by thresholding.	unclear
Yoshida [276]	Nodule enhanced image is produced by computing least asymmetric Daubechies wavelet transform and amplifying responses from intermediate levels before backtransformation. A subtraction image is produced as in [79]. Detected nodules are combined with those detected by [79].	100 images, 122 nodules
Mao [173]	Similar to [13]; for each pixel it is tested if pixels at various distances r display an edge pointed towards the circle.	140 generated nodule images
Wei [259]	Three filters are proposed. They measure if the gradient of edge pixels point to a single center. The edges are located in or on the diameter of a circle, and the diameter and support of these circles is varied.	120 images

Table 2.4: Methods for detecting nodule candidates. The first column lists the first author and reference(s). The method and evaluation data set are described briefly.

study	features	classifier	evaluation
Ballard [13, 15]	Radius, contrast, distance from hilum, shape features.	nearest neighbor	6 images
Sankar [213]	Shape and contrast features from boundary obtained with dynamic programming.	thresholds	8 images, 30 nodules
Sklansky [222]	13 features based on histograms and output of filters applied to ROIs.	thresholds	19 images, 16 abnormal
Lampeter [147, 148]	Rib test (presence of parallel edges), spatial location measures, Hough transform response, radius of circle, gray level features.	linear discriminant	37 abnormal
Giger [79]	Circularity, size and growth rate of region grown area for various thresholds.	thresholds	real and simulated nodules
Giger [80]	Circularity of region grown area for various thresholds after morphological opening.	threshold	60 images, 33 nodules
Cox [51]	Haralick's [101] and Laws' [150] texture features obtained from ROIs with and without nodules.	ANN	ROIs from 5 images
Matsuo [174]	Area after region growing around candidates with different thresholds, degree of circularity and irregularity of regions, magnitude and orientation of single scale gradient within ROIs after background subtraction.	thresholds	60 images, 30 abnormal, 32 nodules
Suzuki [229, 228]	First a non-linear filter, similar to the nodule detector, detects other structures such as vessels, ribs. For remaining candidates, density and circularity features from concentric bands are computed.	thresholds, linear discriminant, feature selection	192 nodules
Peter Chiou [196]	Single scale edge strength and orientation.	Kohonen map	31 images, all abnormal, 87 nodules
Wu [266]	Similar to [174].	ANN, discriminant analysis, thresholds	60 images, 30 abnormal, 32 nodules
Lo [164]	32×32 raw image data from background corrected candidate ROIs.	convolution neural net	55 images, 25 abnormal, 52 nodules

Table 2.5: (continued on next page) An overview of studies that describe methods to classify nodule candidates. The first column lists the first author and reference(s). The second column describes the computed features. The third column lists the classifier where ANN stands for a standard feed forward artificial neural network. The right column describes the database used. This is not necessarily the number of images and nodules used for validation only.

study	features	classifier	evaluation
Lin [158, 159]	32×32 raw image data from background corrected candidate ROIs.	two convolution neural nets	54 images, 23 abnormal [159]
Floyd [72]	Fractal dimension estimated from power spectrum.	threshold	20 images, 10 abnormal
Vittitoe [249]	Fractal dimension estimated from power spectrum.	threshold	30 images, real and simulated nodules
Xu [272]	Diameter, circularity and irregularity measures of candidates obtained by region growing using several thresholds, the slopes of these measures, profile measures, size of regions as a function of thresholds, contrast and gradient measures.	thresholds followed by ANN	200 images, 100 abnormal, 122 nodules
Carreira [33]	Size as a function of thresholds, various shape features from the intensity landscape.	thresholds	35 abnormal images, 70 nodules
Xu [271]	Mean, minimum, maximum, width, standard deviation of histograms of gradients perpendicular to candidate boundaries for various segments of the boundary.	thresholds	200 images, 100 abnormal, 122 nodules
Casaldi [34]	Morphological operations are used to detect likely location of nodules.	–	applied to lossy coding
Nakamura [190]	Computed from outlines of nodules manually determined by radiologists: mean, standard deviation, circularity, ellipticity, irregularity, root-mean-square variation, first moment power spectrum, tangential gradient index, radial gradient index, line enhancement index, mean gradient.	ANN classifies nodules as benign / malignant	56 cases, 34 malignant, 22 benign nodules
Yoshida [275]	After local contralateral subtraction to remove rib structures, signal to noise ratio in candidate ROIs.	threshold	550 candidates, 51 nodules

Table 2.5: (continued from previous page) An overview of studies that describe methods to classify nodule candidates. The first column lists the first author and reference(s). The second column describes the computed features. The third column lists the classifier where ANN stands for a standard feed forward artificial neural network. The right column describes the database used. This is not necessarily the number of images and nodules used for validation only.

The description of features in Table 2.5 is not very detailed. Almost all features are (variations on) “classical features” in the sense that they can be found in many textbooks on image processing, see for instance Chapter 9 in Jain [117]. Classifiers are remarkably often simple thresholds (sometimes of linear combinations of features) but classifiers from pattern recognition theory and standard feed forward neural networks are also used. A few studies focus on the use of a particular type of weakly connected networks [164, 159].

It is clear from Tables 2.4–2.5 that there is considerable overlap between the methods employed in various studies (often by the same researchers; the number of groups that have been working or are working on this problem is limited). Judging the relative merits of methods on the basis of these articles alone seems hardly possible. Comparisons between different methods are rare; sharing databases or setting up common databases for the evaluation of complete systems has not been given much attention. For this particular problem the subtlety of the nodules in the database has a profound impact on detection accuracy and sensitivity, much more than in the case of lung field or even rib cage segmentation. We believe that creating larger databases is key to progress in this field, especially since research on hybrid systems, that combine different methods seems to increase. Unless these databases are publicly available, it will remain very difficult to compare performance. In general, nodule detection remains an open and difficult task in computer-aided diagnosis that is still far from solved. Nevertheless, given its clinical importance, it is likely also to be one of the first areas for commercial exploitation.

The results of the work of Nakamura *et al.* [190] are interesting. Their feed forward neural network outperformed radiologists in classifying nodules as benign or malignant based on features computed from manually drawn outlines of the nodules. Estimating malignancy is a task different from detecting nodules but apparently one in which computer-aided diagnosis could be useful. Gurney and Swensen [92] used a Bayesian classifier and a feed-forward neural network to discriminate benign and malignant nodules using only manually determined features. They reported good results (and the Bayesian classifier outperformed the neural network). Both studies indicate that even if the performance of a computer in *detecting* nodules still lags that of radiologists, they may still serve a role in estimating the likelihood of malignancy once nodules are detected. It may be that computers are better than radiologists at combining the statistical information in large numbers of features, which include both (quantitated!) radiological findings and clinical information. It would be worthwhile to test CAD schemes for nodule classification on larger databases and to investigate if the step of manually outlining can be (partly) automated.

Not mentioned in Tables 2.4–2.5 are a number of observer studies that test the application of CAD schemes for nodule detection [80, 175, 135, 167] or the detection of simulated modules [145], and a paper on synthesis of lung nodules by Sherrier *et al.* [219]. The conclusion of the latter work is that by simply adding nodules that are not obscured by other structures to radiographs at new locations (physically there is nothing wrong with this since radiographs are projection images) it is possible to mimic nodules that cannot be differentiated from real ones by radiologists. This may be very useful to enlarge existing nodule databases.

Texture analysis

The interstitium of the lung is the connective tissue between the blood vessels and the alveoli, the tiny air sacs. There are over 200 types of interstitial lung disease, including pneumoconiosis, sarcoidosis and tuberculosis, and sorting out interstitial disease is one of the most difficult tasks for a chest radiologist [262]. Broadly speaking, one may say that interstitial disease reveals itself as “textural abnormalities”.

The difference between the detection of nodules and that of interstitial abnormalities is, from an image processing point of view, that interstitial abnormalities have a more diffuse character and therefore the two-stage approach characteristic of nodule detection is not easy to apply. Instead of finding candidate abnormalities that can be classified as true or false positives, in principle all areas within the lung fields should be checked for interstitial infiltrates.

In the 1970s, texture analysis in chest radiographs was applied to the detection of pneumoconiosis, or coal miners’ disease. Most of this work used one or more of the by then state-of-the-art methods for texture feature generation, which are reviewed by Haralick [100]. Revesz and Kundel investigated the feasibility of classification based on features computed from the (optically determined!) Fourier spectrum [205]. A similar device was used later by Stark and Lee [226] who compared the performance of several classifiers with a database of 64 chest films. Kruger, Turner, and Thompson [142, 239] computed features derived from co-occurrence matrices (*cf.* [101]) and the Fourier spectrum of manually selected ROIs and classified these regions with linear discriminant analysis. A similar approach was advocated by Sutton and Hall [227]. Later work by Hall [97, 99], used a coarse method for automatic segmentation of the lung fields and presented the design for a complete system. Ledley *et al.* [151] used texture features based on the size and shape of binary versions of the input image obtained by thresholding. They performed classification experiments with 64 films. Jagoe and Patton explored the use of features based on the magnitude and direction of the gradient in several studies [115, 116, 114]. Li *et al.* [152, 153] treated the problem as one of detecting small rounded opacities and employed a strategy comparable to nodule detection. Classification is based on contrast features computed from region grown areas. To our knowledge, none of the work on detecting pneumoconiosis has been used in practice or evaluated for clinical use. This application would be ideal for testing techniques that monitor interval change such as temporal subtraction [124]: the main goal of screening is to assess progression of the disease and coal workers should be screened at least every five years. But no studies have investigated the analysis of radiographs of the same subject taken at different times.

Tully *et al.* [236, 237] were the first to focus on interstitial disease in general, using features from co-occurrence matrices to classify manually selected ROIs with linear discriminant analysis and sequential feature selection.

A large number of studies on the detection of interstitial disease has been performed by Katsuragawa and co-workers. The main elements of their approach are automatic selection of small ROIs within the costal and intercostal space [36], and computation of the standard deviation and the first moment of the Fourier spectrum [126, 127]. These two features are used to distinguish three classes of interstitial disease patterns (nodular, reticular and honeycomb) from normal tissue. The same

method can be applied to the standard radiographs from the International Labour Office for detection of pneumoconiosis [130]. In other studies, geometric features from edges and ridges (line filters) and blobs (found by thresholding) were investigated [183, 129] and improved [112] and horizontal profiles of ROIs were used as input to a feed forward neural network [111]. Features from several systems were combined and both thresholds and neural networks were used for classification. This hybrid scheme obtained a very high accuracy when tested on a database with 200 images from which 100 contained interstitial disease [128, 111]. The use of the scheme was also tested in an observer study [184].

A few studies on the detection of interstitial disease have been done by other groups. Kido *et al.* [133] computed geometric features, similar to the method later described in [129]. In another work, Kido *et al.* [134] classified interstitial abnormalities based on fractal analysis. Some papers deal with the effect of noise and blur in radiographic images on texture analysis methods such as features derived from co-occurrence matrices, the Fourier spectrum, morphological gradients and fractal dimension [245, 246] and try to correct for these effects [185].

In Chapter 8 of this thesis, a scheme is presented in which texture feature vectors for different – overlapping – regions in the lung fields are compared with feature vectors from the same regions in other images. Thus a large number of separate classifiers, one for each region, is used to locally analyze texture. The scheme is applied to a database from a tuberculosis mass screening program and to the database used by Katsuragawa and co-workers. The performance on this database is comparable to the results reported in [111]. The performance on the TB database was substantially lower. This seems to indicate that techniques developed so far are not yet very effective in the detection of subtle cases of interstitial disease. Putting this next to the fact that textural abnormalities may appear in many forms and at all locations, so that it is harder to find candidate abnormal ROIs as in the case of nodules, we may conclude that the detection of subtle interstitial abnormalities is still an open problem.

Other tasks

There are other detection problems in chest radiography that do not fall into the previous sections. An example is the detection of pneumothorax. Sanada *et al.* [212] developed a method to detect pneumothoraces based on a Hough transform technique, evaluated on 50 images with 22 cases of pneumothorax. This is a difficult detection task that could be compared with detecting the tip of catheters [168]. Vogelsang *et al.* [252] present a method to detect catheters but do not present an evaluation.

Conclusions and recommendations

Interpreting a chest radiograph is hard. Even something that sounds as simple as counting the ribs is not easy for an untrained observer. Distinguishing subtle nodules or infiltrates from the normal pattern of branching blood vessels in the lung fields can be difficult even for experienced chest radiologists. When radiologists rate the severity of abnormal findings, large inter-observer and even intra-observer differences occur

[77, 273]. The computer analysis of chest radiographs still presents huge challenges for computer vision researchers, even after thirty years in which over 150 papers on the subject have been published.

Chest radiograph analysis was one of the first subjects to be studied in medical image analysis, but it is no longer an important topic in the field, despite the continuing clinical importance of chest exams. In a recent review concerning 20 years of research on medical image processing, Duncan and Ayache [64] cite work by Sklansky and Ballard [221] as their second reference, but it is also the only reference, out of 139, to work on the analysis of chest radiographs. Within the field of computer-aided diagnosis, mammography currently attracts far more attention and in this field a commercial product is on the market (ImageChecker by R2 Technology, Los Altos, CA). Such products do not yet exist for chest radiography, and there are also no examples of large scale clinical applications of the techniques described in this review.

Many papers report good or excellent results, but this can be deceiving because abnormalities in chest radiographs occur over such a wide range of subtleties. There are several ways to improve the validation and fortunately on average the quality of the evaluation, absent in many early studies, has clearly improved over the years. This can be appreciated by going over the various tables in this chapter where studies are entered in chronological order. Segmentation methods should preferably include quantitative results, offset against intra- and inter-observer variability. The true value of schemes that detect abnormalities can only be assessed when they are compared in observer studies with the performance of radiologists, or if the performance of radiologists with and without the aid of a computer are compared. In the latter case the task for the radiologist should be similar to clinical practice.

More importantly, databases should contain a sufficient number of cases and should at least display the normal variety encountered in clinical practice. Publicly available databases could provide a great impetus to the quality of validations. Such data could also make it possible to benchmark different methods. Making code implementations publicly available would facilitate comparative studies. Such investigations have hardly been made in the existing literature. Increasing the size of testing databases is also likely a key factor to improvements in CAD schemes. Often the number of actually positive cases is rather limited compared to the number of features used for classification or feature selection or the number of parameters used in rule-based schemes. Duncan and Ayache [64] emphasize the importance to develop appropriate validation approaches, together with the need to set up test databases for the research community, for the whole field of medical image processing.

Another recommendation from [64] that seems directly applicable to computer analysis of chest radiography is the need to think more seriously about core principles and techniques that drive applications in medical image analysis. This may prevent the field to “re-invent the wheel” ever so often. Duncan and Ayache note that the bulk of efforts over the last decades consist of identifying separate tasks and applying engineering principles to design appropriate solutions. Indeed, almost all literature discussed here follows this somewhat ad hoc approach. Applying more general techniques or developing methods that employ the physics of the image formation process of chest radiographs are rare. Overall, the big challenges for future research on the analysis of chest radiographs are similar to those of other fields in medical imaging.

There is a need to incorporate knowledge in the algorithms in an effective way, to perform global instead of local analysis, and to successfully integrate different kinds of information.

We end this review by identifying a number of trends in the field. Some of them are already evident from the literature, others are our personal recommendations for further research:

- Focus on solving tasks that are encountered in clinical practice. Use data that contain abnormalities as they are encountered in practice, *cf.* the list given by MacMahon ([168], Table 2.3).
- Use larger databases for better validation. Make databases and/or code implementations publicly available, to facilitate comparative studies and to reduce the burden of individual research groups for collecting data. To increase the size of databases, it can be useful to explore the use of methods that use existing data to simulate new cases.
- The number of studies in which several methods are combined is increasing. Hybrid systems that use several types of features, sometimes including clinical information, are used to eliminate false positive nodules or to diagnose diseases.
- The number of radiology departments that use central systems to store digital chest radiographs is increasing. This provides an opportunity to use “chest workstations” with intelligent algorithms that run in the background and may alert radiologists to possible abnormalities. This requires a more integrated approach to processing radiographs as they are encountered in clinical practice, while almost all studies published so far focus on a single aspect or abnormality to be detected.
- Subtraction techniques are becoming a powerful tool to enhance the visibility of abnormal structures. There is room for improvement in knowledge-based approaches that use the statistical data from large numbers of images to model the normal structures in a chest radiograph. The use of dual energy subtraction images (for an example see Figure 9.1) in computer analysis, that can be produced virtually for free by many chest units, could improve the sensitivity of CAD systems.
- Accurate segmentation of bony structures such as the rib cage, including posterior and anterior ribs, and the clavicles, of branching blood vessels, and of objects such as clothing and catheters remains an open problem. Such segmentations could be used in analyses to eliminate false positives, to choose regions of interest for texture analysis and to subtract normal structures.
- We expect to see more applications of knowledge based techniques that are based on image statistics. Such techniques, *e.g.* segmentation with active shape models, are gaining popularity in many parts of medical image processing and the analysis of chest radiographs, where large numbers of images are available, seems ideally suited for their application.

- So far, most research efforts in CAD have concentrated on the detection of possibly malignant structures, but recently the classification of abnormal signs is getting more attention. Experiments in which nodules are categorized as benign or malignant or in which signs of interstitial disease together with clinical findings are automatically diagnosed, have shown very promising results. Therefore we expect that computer algorithms may become an important aid to the radiologist in differential diagnosis.

Chapter 3

Segmenting the Lung Fields

Nine algorithms for the automatic delineation of lung fields in standard PA chest radiographs are presented. The algorithms are based on different techniques: matching, pixel classification based on several combinations of features, a rule-based scheme that finds lung contours using a general framework for the detection of oriented edges and ridges in images. Each approach is discussed and the performance of the nine systems on 115 test images is compared with inter-observer variability and results available from the literature. Best performance is obtained by a hybrid scheme that combines the rule-based segmentation algorithm with a pixel classification approach. This combination of two complementary techniques leads to robust performance: the accuracy is above 94% for all images in the test set. The average accuracy of the scheme is 0.969 ± 0.008 , which is close to the inter-observer variability of 0.984 ± 0.005 .

CHEST RADIOGRAPHS are of paramount importance in the identification of patients with abnormal pulmonary conditions. A plain chest radiograph frequently also provides a general indication of the type of pathology present. The advent of digital thorax units [74] and digital radiology departments with PACS makes it in principle possible to use computerized methods for the analysis of chest radiographs on a routine basis. In this work, we focus on automated segmentation of lung fields in standard PA chest radiographs. This is a mandatory pre-processing step for most types of computer analysis of chest radiographs. The subject has received a considerable amount of attention in recent literature (see also Chapter 2). The two main approaches are rule-based reasoning and pixel classification with neural networks. With rule-based systems we mean algorithms consisting of a series of steps, each containing specific processing and, usually, certain adjustable parameters. For the segmentation of lung fields, such schemes have been proposed by Xu *et al.* [269,270], Duryea and Boone [65], and Carrascal *et al.* [32]. Lung segmentation by pixel classification using neural networks has been investigated by McNitt-Gray [179,178], Hasegawa *et al.* [103], and Tsujii *et al.* [235]. Vittitoe *et al.* [250], [251] developed a pixel classifier for the identification of lung regions using Markov random

field modeling. Brown *et al.* [28] presented a system to extract lung edges that employs reasoning mechanisms. A third possible approach is to match an input image to segmented reference images. Although matching to atlases has been applied to several problems in medical image processing, the direct application to lung field segmentation has not been explored before.

The fact that statistical classifiers and rule-based schemes seem to be the most popular methods for thorax segmentation is perhaps not surprising because knowledge-based processing is mandatory to solve this task. There are several reasons why the automatic segmentation of chest radiographs is a hard problem from a computer vision point of view. First, there are large anatomical variations in the chest from person to person. Second, the habitus and level of inspiration of the subject during the examination have a profound impact on the image. Third, the settings of the chest unit, particularly the peak tube voltage, determine how well bony structures are visible [10]. But finally, and most importantly, radiographs are projection images and thus contain superimposed structures. Physicians are actually trained to *mentally subtract* anatomical structures selectively. To cite one popular textbook [225]: “The system generally employed by the radiologist is to *look at* various structures in a deliberate order, concentrating on the anatomy of each while excluding the superimposed shadows of other structures.” The lung fields in chest radiographs contain several superimposed structures, such as lung vasculature, posterior and anterior ribs, and clavicles. These structures do not make up the borders of the lung fields, as opposed to other structures, such as the mediastinum and the diaphragm. Analysis should differentiate between structures and this can only be done by the incorporation of knowledge.

We consider both a rule-based system and pixel classification with several sets of features. The rule-based scheme we present is new and uses a general and flexible framework for the detection of subdimensional structures, in this case lung edges, about which knowledge is available. Our pixel classifiers use features such as intensity and location, but also features derived from the rule-based analysis and an entropy measure. Furthermore, we propose a new hybrid segmentation scheme that combines the strength of a rule-based approach and pixel classification. We show that this system comes close to the accuracy of segmentation by hand and compares favorably with literature results. Such direct comparisons between different segmentation methods have not been made before.

In the next section we define the exact task and discuss the database we used in our experiments. Subsequently, we will consider each of the three approaches proposed, discuss their advantages and disadvantages and describe the algorithms in detail. In the final sections we present results, discussion and conclusions.

Purpose, materials and methods

A possible definition of lung fields is those parts of the image for which the radiation has passed through the lungs. However, this *total projected area* is impossible to determine from a chest radiograph because of the density of the overlying diaphragm, heart and mediastinum. Therefore, we define the lung fields as those parts of a chest

radiograph which contain lungs not obscured by either diaphragm, mediastinum and heart. It has been estimated [39] that on average 26% of the lung volume and 43% of the total projected area is obscured by one of these structures.

Our database consists of standard PA chest radiographs taken in a tuberculosis screening program for people seeking political asylum in The Netherlands. These images are read by two physicians independently and classified as normal or possibly abnormal. In the latter case, the patient is contacted for further examination. We randomly selected 133 normal and 133 abnormal cases. The images were taken with a mobile Oldelft Electrodelca (Nucletron BV, Veenendaal, The Netherlands), a system commonly used in mass chest screening. The tube voltage was 117 kV and the images were printed on 10 by 10 cm. film. The films were digitized with a Lumisys 100 scanner (Lumisys, Inc., Sunnyvale, CA) to 996 by 996 pixels with 10 bit intensity, subsampled to 256 by 256 pixels, which is sufficient for our purposes. We excluded images from subjects less than 16 years old, and images for which our rule-based segmentation scheme failed to produce any output, indicating gross abnormalities. In these images, 5 in total, at least one of the lung fields was not visible. The remaining 230 images were randomly divided in a test set and a training set, each containing 115 images, about half of which were abnormal. Note that this percentage of abnormal images is much higher than what is encountered in mass chest screening practice. Nevertheless, we did not want to exclude abnormal cases because robust segmentation methods should give good performance on abnormal images as well. Figure 3.1 displays thumbnails of the images to give an idea of the variation encountered in the database.

We indicate the coordinate of the upper left corner of the image with $(0,0)$, and the lower right with $(1,1)$. All references to coordinates and lengths are made in these units, which is convenient because it makes them independent of the image size in pixels. The lung fields were traced with a mouse by two observers (the author (BvG) and his co-worker Bart ter Haar Romeny (BtHR)) independently, under supervision of an experienced radiologist. BvG traced the contours in both the training and the test sets,; these contours are used as the gold standard in the evaluation. BtHR traced all contours in the test set. These contours were used to compute the inter-observer variability.

Approaches to lung segmentation

This section discusses the various possible approaches to lung segmentation.

Matching

Matching, or registration, of images can be defined as finding a transformation that relates points in one image to their corresponding points in another image. Matching can be used for segmentation by matching a segmented reference image to an input image. The transformed reference image yields a transformed segmentation that, if the images are well matched, will be an accurate segmentation of the input image.

A general matching system is determined by four elements [27]. First of all, there is the choice of feature space. This amounts to selecting a reference image, deciding

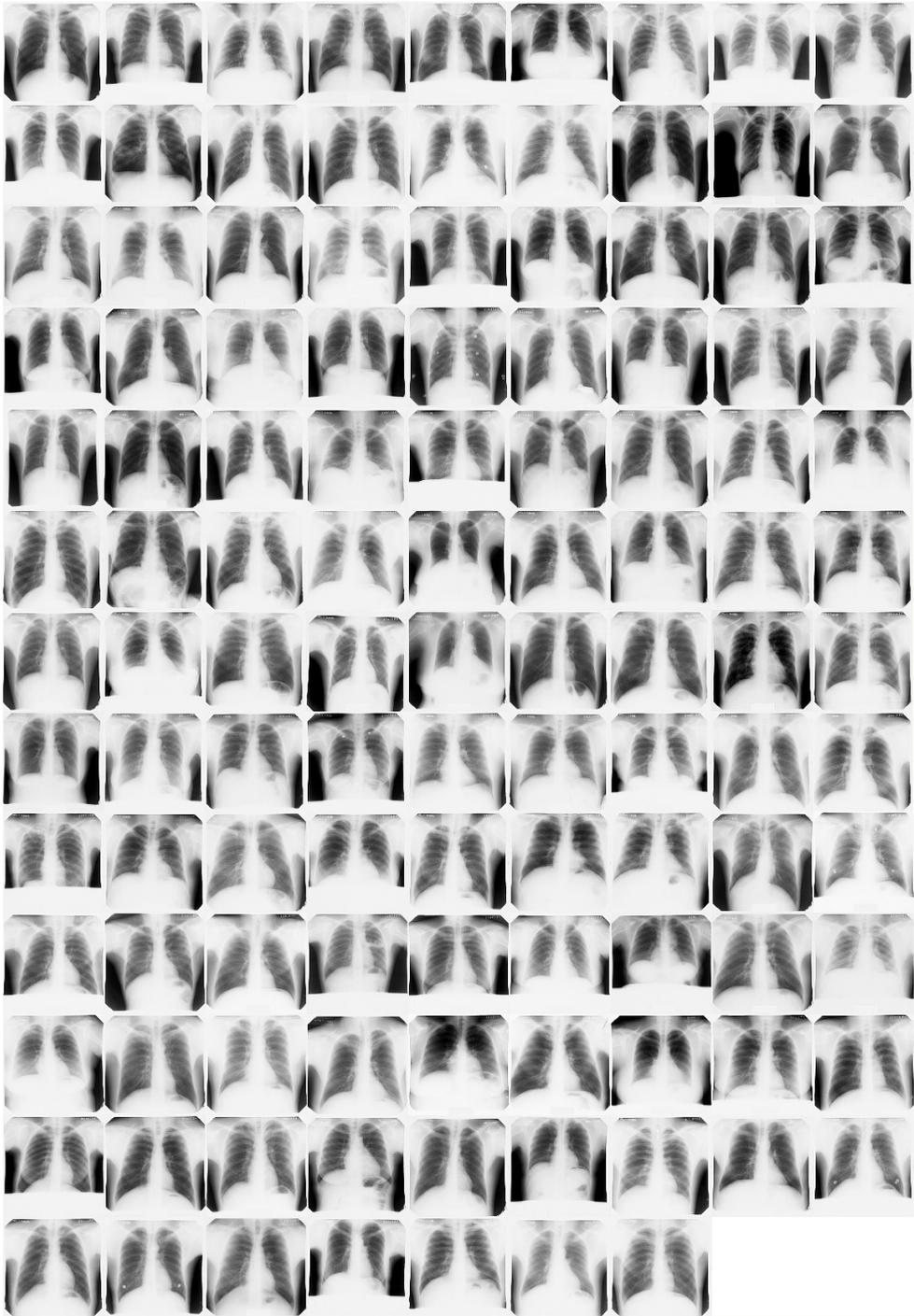


Figure 3.1: The 115 test images on which segmentation algorithms are evaluated.

on which points in this image to use and possibly applying certain operations on these points, such as filtering. Clearly the choice of a reference image is an important one since the anatomical variation between chest radiographs is large and the selected reference image is the model used during matching. If the reference image is not similar to the input image, up to the transformation searched for, matching will fail. Second, the search space is defined by the parameter space of the allowed transformations. In the case of lung segmentation using reference and input images from different subjects, it is evident that transformations that allow anisotropic scaling are required. One has to choose between global transformations, such as an affine transformation and local, elastic, deformations. Third, a similarity metric is required that should indicate how well the transformed input image and the reference image match. The final element is the search strategy, *i.e.* the algorithm used to determine the transformation which maximizes the similarity metric. Details about the system used are given in the next section.

An important advantage of this approach is that it incorporates model knowledge and shape information in an implicit way. There is no need for specific analytical models of the objects to be segmented, the reference image *is* the model.

Rule-based detection of lung contours

Rule-based systems offer their designer the freedom to express his knowledge about the problem in any type of rule or processing imaginable. Furthermore, it naturally subdivides the problem in sub-problems. If a part of the system does not yield satisfactory performance, one can add pre- or post-processing steps to correct the problem. There are some obvious weaknesses as well. In practice, rule-based systems consist of a concatenation of steps, each containing several variables, thus leading to an overall scheme that often contains many ad hoc choices and a myriad of user-adjustable parameters. It is usually impractical or impossible to make plausible, let alone prove, that the performance of the system is in some sense optimal. The natural way to improve the system is adding more rules, thus also adding complexity. The same system cannot be applied to a different task: rule-based systems do not generalize.

Another pitfall is the fact that these systems usually take the outcome of one step as input for the next. This may lead to performance that is not robust. In practice, such systems may perform reasonably well for each substep considered separately, but go astray when applied as a whole.

The starting point of our rule-based approach is the observation that the borders between anatomical structures in chest radiographs often largely coincide with edges and ridges in the image. An illustration is given in Figure 3.2. A difficulty is that these structures correspond only partly to the borders between anatomical regions. It is not straightforward to pick out the “correct” structures or parts of structures. Note for instance in the top-right image of Figure 3.2 that the edge between right lung and mediastinum is connected to the edge of the clavicle. We propose a dimensionality reduction technique to overcome some of these problems.

Edges are usually defined as those points in an image where the gradient magnitude is maximum in the gradient direction [31]. Ridges are commonly defined as

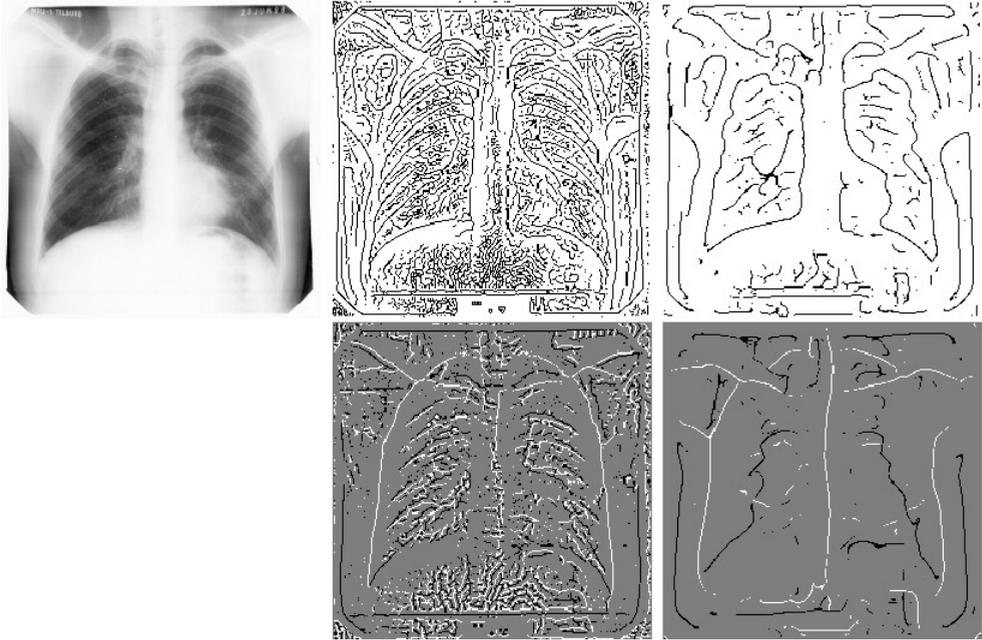


Figure 3.2: Edges and ridges in chest images coincide with borders between anatomical structures. A chest radiograph (top-left) with (top-middle) edges at a fine scale of 0.004 and (top-right) a coarser scale of 0.015 and (bottom-middle) ridges (in white) and valleys (in black) at the same fine and (bottom-right) coarse scale. Scale is expressed in image-width units.

extrema in the direction of the largest curvature. Instead of these definitions, we consider structures defined by extrema in a *fixed* direction, for derivatives of a certain order in that same direction.

Derivatives are computed by convoluting the image $L(x, y)$ with the derivative of a Gaussian $G(x, y; \sigma)$ at a particular scale σ . The normalized Gaussian in 2-D is given by

$$G(x, y; \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(\frac{-(x^2 + y^2)}{2\sigma^2}\right). \quad (3.1)$$

If we denote the n th order derivative in the direction defined by an angle α as L_n^α we need to compute

$$L_n^\alpha(x, y; \sigma) = G_n^\alpha(x, y; \sigma) \otimes L(x, y), \quad (3.2)$$

where \otimes denotes convolution and G_n^α is the n -th order derivative of the Gaussian kernel in the direction α . Instead of calculating the convolution directly, it is often convenient to express the response for an arbitrary direction α in terms of combinations of a

finite set of basis filters (such constructions are commonly referred to as *steerable* filters). In particular in 2D, the derivatives of any order n for any direction α can be computed from the set of all derivatives up to order n derivatives in two fixed orthogonal directions, e.g. x and y by substituting $x' \rightarrow x \cos \alpha - y \sin \alpha$ and $y' \rightarrow -x \sin \alpha + y \cos \alpha$. Other basis filters are also possible. See [73] for a general discussion on steerable filters. In particular, for order $n = 1$ and $n = 2$ one obtains

$$L_1^\alpha(x, y; \sigma) = \cos \alpha L_x + \sin \alpha L_y, \quad (3.3)$$

$$L_2^\alpha(x, y; \sigma) = \cos^2 \alpha L_{xx} + \cos \alpha \sin \alpha L_{xy} + \sin^2 \alpha L_{yy}, \quad (3.4)$$

where on the right-hand side subscripts denote differentiation to x and y . The detection of extrema is simply performed by comparing each pixel (x, y) with two neighboring locations $(x + \eta \cos \alpha, y + \eta \sin \alpha)$ and $(x - \eta \cos \alpha, y - \eta \sin \alpha)$. If both locations, are lower/higher than the given point, it is a maximum/minimum. We fix η at 0.004 and use linear interpolation.

In this way we obtain structures that cannot cross, because in the direction defined by α consecutive points cannot both be extreme. This implies that such structures can never “turn”, *i.e.* a line in the direction defined by α cannot cross a structure more than once. The order of derivative n and the choice for minima and maxima determines the nature of the detected structures. Using order $n = 0$, one detects the axes of bright and dark blobs; $n = 1$ yields edges from dark to bright regions, or vice versa; $n = 2$ finds bright and dark line structures, all in the direction perpendicular to α . A significance measure for a structure can be computed by integrating $|L_n^\alpha|$ along the structure or by taking the accumulated length of a structure.

Some results are shown in Figure 3.3. Using the appropriate scales and combinations of directions and derivatives, we extract structures that correspond closely to anatomical structures or boundaries, and which are not connected, such as the structures in Figure 3.2. In Figure 3.3 the thorax center line and the lung center lines are extracted (top-left), the edges of the mediastinum (top-middle), the rib cage (top-right), the rib borders (bottom-middle), and the centerline of costal and intercostal spaces (bottom-right).

The problem of connected structures has greatly been reduced in this way. Since the orientation of lung fields is approximately the same for each radiograph, we can use this knowledge to select the proper directions. Another tunable parameter is the scale σ at which derivatives are computed. In general, scale selection is an active research area, both for single scale and multi-scale methods. In this study we used empirically selected values. One rule of thumb should be that the scale corresponds to the size of the structure to be detected. In practice it may be worthwhile to consider coarser scales since this leaves less structures, which may alleviate the selection of the “correct” structure. To solve this selection problem, we use a straightforward voting technique, the details of which are explained in the algorithm in the next section.

For the detection of diaphragm and lung top, one could use the same techniques. However, the simple voting mechanisms may fail to detect the correct structure. This is caused by, as can be verified from the bottom row of Figure 3.3, the presence of several horizontal edges around the diaphragm and horizontal or circular bright line

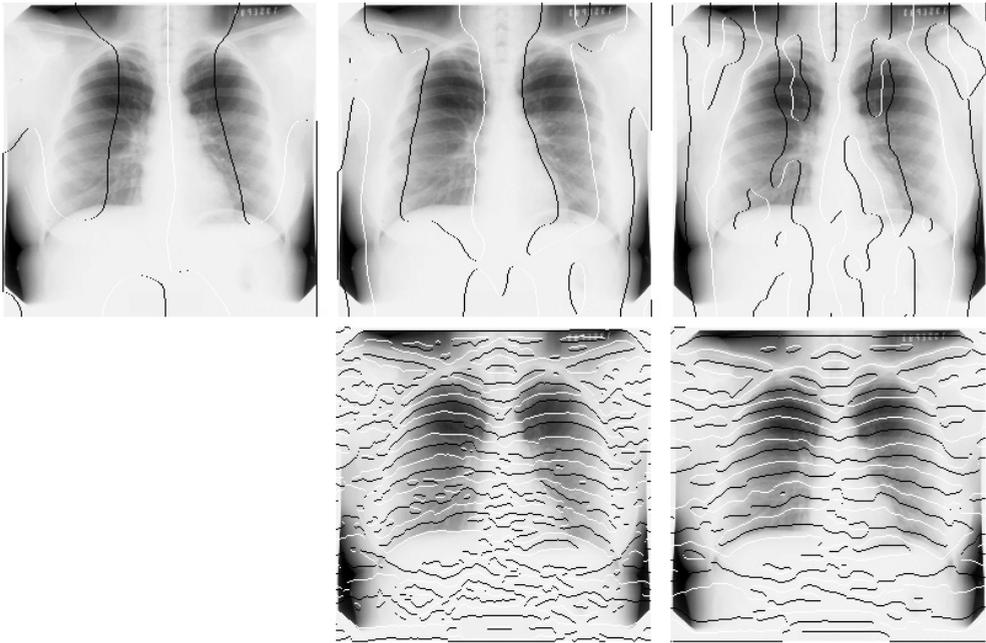


Figure 3.3: Oriented extrema coincide with borders between anatomical structures. A chest radiograph with superimposed (top-left) extrema (maxima in white, minima in black) in the x -direction of the 0th order derivative at a scale of 0.047; (top-middle) extrema in the x -direction of the 1st order derivative at a scale of 0.039; (top-right) extrema in the x -direction of the 2nd order derivative at a scale of 0.031; (bottom-middle) extrema in the y -direction of the 1st order derivative at a scale of 0.0078; (bottom-right) extrema in the y -direction of the 2nd order derivative at a scale of 0.016.

structures in the neighborhood of the lung top. Since the lung top and diaphragm are generally the strongest lines and edges, we used dynamic programming [18, 3] to detect them. In the algorithm outlined in the next section we will use the fact that the diaphragm corresponds to a maximum in the 1st derivative in the y -direction and that the top of the rib cage corresponds to a maximum in the second derivative perpendicular to the rib cage, together with the fact that the contour corresponding to the top of both lungs is approximately circular. Putting this together, estimates of the diaphragm, the border between lung fields and mediastinum, the boundaries of the rib cage and the lung tops are obtained. Figure 3.4 shows representative results.

This algorithm does not necessarily produce the complete contours of both lung fields, since it is not enforced that there are crossings between the mediastinum edge, the lung top, the rib cage, and the diaphragm respectively. In some cases, (parts of) these contours may be missing. We consider this to be a strength of the method, because the fact that structures are not detected may indicate abnormalities. One example is shown in Figure 3.4(a). In fact, this was used to eliminate 5 grossly abnormal images from the database, as mentioned in the previous section. Forcing

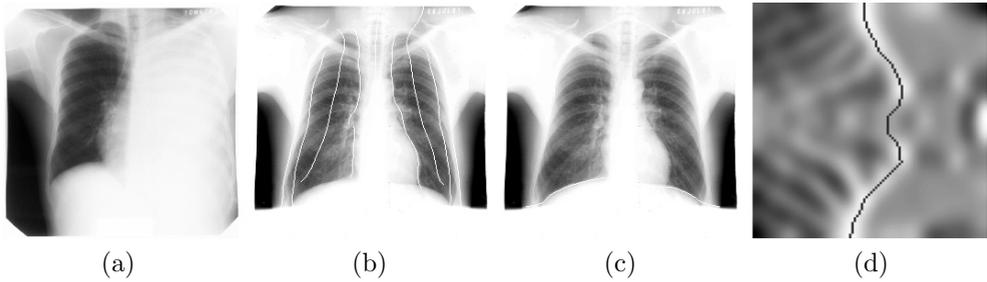


Figure 3.4: (a) Example of an image that is rejected because no thorax center line is found; (b) Typical result of detection of vertical structures (thorax center line, mediastinum edges, lung edges, rib cage); (c) Typical result of detected diaphragm and lung top; (d) Polar transformation of a hemi-circle of the image used in the detection of the lung top. The horizontal axis is the radius, the vertical is the polar angle.

the detection of structures by fitting models to the data might leave abnormalities unnoticed.

The results of this algorithm are converted to a complete lung contour using a matching approach. We take a lung contour from a reference image and select control points along this contour. These points are moved in horizontal or vertical direction until they coincide with the detected mediastinum, lung top, rib cage or diaphragm respectively. The transformed contours, based on the displacements of control points, define the lung fields. The detected structures are also used to obtain a set of landmark points for each image. These landmarks give an estimate of the width, the height and the location of the center of each lung field and are used to compute features in pixel classification schemes.

Pixel classification

Segmentation can be treated as a pixel classification problem by calculating a feature vector for each pixel in the input image. Output is the anatomical class the pixel belongs to. Any classifier from statistical pattern recognition or neural network theory can be used to approximate this mapping. The classifier is trained with a large set of training samples (pixels from a large collection of training images). Although different types of classifiers will obviously lead to different results, the performance of these segmentation algorithms will depend mostly on the features of the input vector.

One of the main strongholds of this approach is the fact that the results are guaranteed to be optimal in some sense; there are no parameters to be set by a user as compared to the rule-based approach. Consider a simple classifier that only uses intensity as a feature; if the training data is representative of the test data, the classifier will choose the optimal (even multiple) threshold, as opposed to a rule-based thresholding scheme. Another attractive element is the fact that knowledge is implicitly integrated in this approach through the use of training data. Any “general” segmentation technique (like region growing techniques or watershed methods) needs

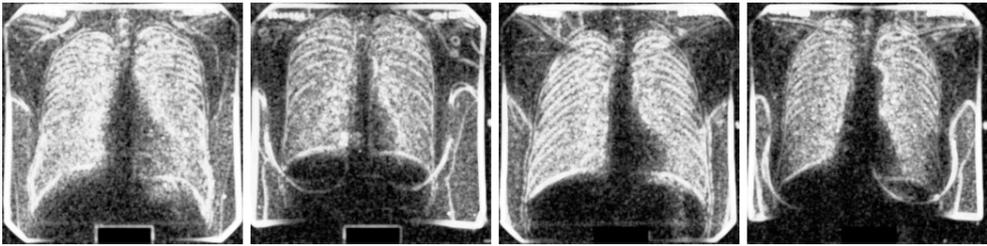


Figure 3.5: The entropy measure, discussed in the text, applied to 4 chest radiographs, giving high output in the lungs and at edges outside the lungs. The images were equalized for display purposes.

to be fine-tuned for a particular application, which involves the choice of parameters. In this approach, training the classifiers leads to optimal parameters. Note however, that the actual choice of feature set (and the parameters used to compute them) is, in general, completely arbitrary and a choice of the designer of the system.

In most implementations, pixel classifiers use local features only. It is important to realize the limitations of such systems. It seems impossible, even for human observers, to always correctly classify a very small region as lung or non-lung without the *global* percept of the whole image. Pixel classification approaches have been characterized as region-based and consider rule-based systems as contour-based. This is not necessarily true. One can use contour-based features, such as the distance from a location to a strong edge, in a pixel classification approach. In a rule-based scheme, region measures such as average intensity over a region of interest may be used.

Let us now consider suitable candidate features. The lung fields have a typical shape and location in the image. This suggests pixel location as a feature. The intensity is a good feature since pixels inside the lung fields have lower values than those of the surrounding tissues but higher values than areas outside the body. Furthermore, the lung fields contain overlying ribs and lung vasculature. This yields local variations in intensity at a small scale, while areas outside the lungs are often more homogeneous. We have formalized this notion as follows: subtract a slightly blurred version of the image from the original, thus obtaining a narrow high-pass frequency filtered version of the image; take the absolute value of this image and blur it slightly. This procedure contains two scale parameters and can be considered an entropy measure since it has high response in regions with large local variations. Figure 3.5 shows that this operation indeed produces high output in the lungs, but also at the borders of the film. Finally, the rule-based scheme detects the rib cage, mediastinum edges, diaphragm and lung tops. From the structures detected by the rule-based scheme, a transformation can be computed that scales the width and height of the lungs to standard values (scaling in x - and y -direction) and translates the thorax with its center to a standard location. This 4 parameter transformation allows the computation of a *corrected* location for each pixel.

Hybrid scheme

One could argue that pixel classification methods are in some sense complementary to the rule-based method we have discussed previously. This leads to the question whether it is possible to combine both methods. We present a simple method to do this. From all images in the training set, we determine two training sets consisting of all pixels which the rule-based method classified as lung and the pixel classification method as background, and vice versa. We use the corrected location as feature. These sets can be used for reclassification. This hybrid system (System 9) is the most advanced scheme we consider in this paper.

Algorithms for lung segmentation

In this section we will give a detailed description of 9 segmentation schemes that are evaluated in the subsequent experiments.

1. Matching

We selected one (qualitatively judged as representative) normal thorax as reference. As feature space we used the gray level intensity of the pixels of the image, normalized to zero mean and unit standard deviation and subsampled to 64 by 64 pixels. We allowed affine transformations and used the sum of absolute differences in intensities as similarity metric. We used Powell's direction set method [202] to find the optimal transformation.

2. Rule-based reasoning

- (a) Detection of thorax center line. The ROI was a rectangular area defined by $0.3 < x < 0.7$ and $0.2 < y < 0.6$. This range for y was also used in steps (b)-(e). The 0th order derivative was computed at a scale of 0.05. Pixels that are maximal in the horizontal direction were detected and grouped into 8-connected structures. For each horizontal line in the ROI, the structure closest to the vertical centerline of the ROI was voted for. The structure that received most votes was selected. Other structures were added, in order of received votes, as long as they did not overlap in their y -coordinates. If no center line was detected, the image was rejected.
- (b) Detection of the right/left edge of mediastinum. The ROI was bounded on the right/left by the thorax center line and has a 0.4 width. The 1st derivative in the x -direction was computed at a scale of 0.03. Maxima/minima in the horizontal direction were grouped into connected structures. For each horizontal line, starting from the thorax center line, the first structure encountered was voted for. Again, other structures were added, in order of received votes, as long as they don't overlap in their y -coordinates. The same procedure is used in steps (c) to (e).
- (c) Detection of lung center lines. The mediastinum edges are the right and left boundaries of the ROI, which had a 0.3 width. Horizontal minima of the 0th order derivative at a scale of 0.03 were used.

- (d) Detection of right/left lung edges. The lung center lines are the right and left boundaries of the ROI which has a 0.4 width. Horizontal minima/maxima of the 1st order derivative in the x -direction at a scale of 0.03 were used.
- (e) Detection of right/left rib cage. The lung edges are the right/left boundaries of the ROI which has a 0.2 width. Horizontal maxima of the 2nd order derivative at a scale of 0.02 were used.
- (f) Detection of diaphragm. The right and left diaphragm were determined with dynamic programming which finds a line from left to right in a rectangular ROI. The ROI contained the 1st derivative in the y -direction of the image at a scale of 0.02. The line was allowed to have a slope between -1 and 1. The right diaphragm was detected first, because it is visible more clearly, due to the absence of the heart and stomach bubbles on the left side of the image. The x -range of the ROI was from the lowest x -coordinate of the rib cage to the lowest x -coordinate of the thorax center line. A starting point on the right diaphragm was detected by finding the right lung center at $y = 0.4$ and taking a vertical profile from this point and finding the highest value for the first derivative along this profile. The y -range of the diaphragm was 0.4, centered around this point. After the right diaphragm had been detected, crossings with the right rib cage and mediastinum edge were determined. These landmark points were added to the right diaphragm starting point. This resulted in 1, 2, or 3 landmark points. The y -range of the ROI for the left diaphragm was given by the smallest y -coordinate of these points minus 0.2 of the image height and the largest y -coordinate plus 0.2 times the image height. The x -range of this ROI was from the highest x -coordinate of the thorax center line to the highest x -coordinate of the left rib cage.
- (g) Detection of lung tops. We used the fact that the top is a bright line structure, just like the rib cage. It is more or less circular, so a polar transformation is made of a hemi-circle in the image to 100 by 100 pixels. As center point the thorax center line at $y = 0.4$ was taken. The radius of the hemi-circle was estimated to be half the distance between the right and left rib cage at this height. The minimum and maximum radius were set at 0.7 and 1.3 times this radius respectively. We took the 2nd derivative along the circle radius and found a maximum path with dynamic programming. In the maximum cost path, we allow deviations from a circular path of at most 45° .
- (h) Lung contours from a reference image with 15 control points along each contour were moved in horizontal or vertical direction until they coincide with the detected mediastinum, lung top, rib cage or diaphragm respectively. The contour interpolated based on the displacements of control points was taken as final lung contour.
- (i) Landmark points at half the height and width of the lungs were determined for later use.

3. Pixel classification: most likely class

This is a trivial classifier in which each pixel was assigned to the most likely class, the background.

4. Pixel classification: location

In the systems 4 to 9 we used a k -nearest-neighbor classifier with $k = 31$. This setting of k was determined as optimal in several pilot experiments. We used 1024 points from each image in the training set (32 by 32). The classification was done for each image in the test set, subsampled to 256 by 256 pixels. In all systems, scaling factors were calculated to obtain zero mean and unit standard deviation for each feature over the whole training set. These scaling factors were also applied to the features of pixels in the test images.

In system 4 we only used the pixel coordinates as a feature. This yields a classifier that is independent of the input image.

5. Pixel classification: intensity

In this system we used only pixel intensity as a feature. This results in a system that segments the lung fields by an optimal (multiple) threshold.

6. Pixel classification: intensity and location

Both intensity and the coordinates of each pixel were used as features.

7. Pixel classification, rule-based reasoning: intensity and corrected location

Using the landmarks obtained from the rule-based system, we computed a horizontal and vertical scaling factor that scales the estimated width and height of the right lung to fixed values of 0.235 and 0.547. After that, we calculated the translation that takes the center of the thorax to (0.5,0.5). This yielded a corrected location for each pixel that was used as two features (x - and y -coordinates), together with its intensity.

8. Pixel classification, rule-based reasoning: intensity, entropy and corrected location

This system was equal to System 6, with an entropy measure added as fourth feature, in addition to intensity and corrected pixel coordinates. The entropy was defined by blurring the image to a fine scale of 0.0039, subtracting this blurred image from the original, taking the absolute value and blurring the result to a scale of 0.0039.

9. Rule-based reasoning and correction using pixel classification

This system reclassified those pixels in the image for which systems 2 and 7 yielded different classes. There are two possibilities; pixels can be classified as lung by system 2 and as non-lung by system 7, or vice versa. For each case, we determined a training set consisting of all the pixels in the training set for which this situation occurred (subsampled to 256 by 256). We used the corrected location as features.

Results

Segmentation performance can be measured in various ways, and ultimately, the requirements of the application that uses the segmentation as input determine whether the segmentation is sufficiently accurate. Here we consider the problem as segmentation between lung and background and calculate the classical accuracy, sensitivity and specificity

$$\text{Accuracy} = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{TN} + N_{FP} + N_{FN}}, \quad (3.5)$$

$$\text{Sensitivity} = \frac{N_{TP}}{N_{TP} + N_{FN}}, \quad (3.6)$$

$$\text{Specificity} = \frac{N_{TN}}{N_{TN} + N_{FP}}, \quad (3.7)$$

to measure segmentation performance. N_{TP} is the true positive fraction (part of the image correctly classified as lung), N_{TN} is the true negative fraction (part of the image correctly classified as background), N_{FP} is the false positive fraction (part of the image incorrectly classified as lung), and N_{FN} is the false negative fraction (part of the image incorrectly classified as background).

The contours manually traced by BvG were used as gold standard. The lung contours in the test set were manually segmented by BtHR, and from these contours the inter-observer variability was computed.

In the literature, other methods are used to evaluate the performance of segmentation schemes. One popular approach is to present the results to radiologists who rate the performance on a qualitative scale. A drawback of this method is that it is impossible to compare results of such evaluations with other studies. Accuracy, sensitivity and specificity have been used, or can be calculated from results reported, in several literature studies.

Table 3.1 lists the results of our schemes together with results reported in the literature. The results are sorted according to accuracy. Key result is the accuracy of 96.9% of system 9, which approaches the inter-observer variability of 98.4%. Another interesting result is the fact that we did not find significant differences in segmentation accuracy for normal versus abnormal radiographs.

Figure 3.6 shows results of several systems for 5 images. The images were chosen to range from best to worst performance for System 9, the overall most accurate system. Figure 3.7 shows the fraction of false positives and false negatives as a function of location for systems 8, 2 and 9.

All experiments were performed on a standard PC with a 450 MHz Pentium III processor. The rule-based segmentation (system 2) required 4.1 seconds on average. Pixel classification schemes required anywhere from 8 seconds to 4 seconds. The reclassification in system 9 requires the computation time of both the rule-based scheme, the pixel classifier of system 7, and reclassification. In total, this took 10.4 seconds on average. Segmentation by matching took about 24 seconds.

Method	Accuracy	Sensitivity	Specificity
Inter-observer variability	0.984 ± 0.00475	0.957 ± 0.0174	0.993 ± 0.00306
Classification correction (9)	0.969 ± 0.00803	0.943 ± 0.0330	0.978 ± 0.0106
Rule-based (2)	0.961 ± 0.0116	0.940 ± 0.0389	0.969 ± 0.0153
PC int., entropy, corr. location (8)	0.956 ± 0.0157	0.912 ± 0.0617	0.972 ± 0.0248
PC int., corrected location (7)	0.953 ± 0.0177	0.906 ± 0.0649	0.970 ± 0.0288
PC int. and location (6)	0.933 ± 0.0219	0.854 ± 0.0771	0.966 ± 0.0309
Matching (1)	0.907 ± 0.0437	0.766 ± 0.153	0.946 ± 0.0480
PC location (4)	0.898 ± 0.0382	0.784 ± 0.0984	0.947 ± 0.0339
PC int. (5)	0.847 ± 0.0356	0.727 ± 0.122	0.891 ± 0.0447
All negative (3)	0.736 ± 0.0551	0 ± 0	1 ± 0
Duryea: rule-based method [65]	0.959 ± 0.054	0.863 ± 0.11	0.987 ± 0.044
Vittitoe: MRF [250]	0.948 ± 0.016	0.907 ± 0.044	0.972 ± 0.020
McNitt-Gray: 59 features [178]	0.932	0.949	0.922
Tsujii: PC corr. int., loc. [235]	0.923		
McNitt-Gray: 8 features [178]	0.918	0.903	0.930
Vittitoe: PC int. 3x3 [250]	0.893 ± 0.027	0.846 ± 0.057	0.925 ± 0.038
Vittitoe: PC location [250]	0.880 ± 0.035	0.820 ± 0.08	0.920 ± 0.044
Duryea: classifying location [65]	0.879 ± 0.063	0.785 ± 0.10	0.934 ± 0.063
Vittitoe: fixed thresholding [250]	0.806 ± 0.071	0.860 ± 0.058	0.781 ± 0.112
Duryea: all negative [65]	0.751 ± 0.063	0 ± 0	1 ± 0

Table 3.1: Accuracy, sensitivity and specificity of all 9 systems on the complete test set. PC stands for pixel classification, MRF for Markov random field model. The number in parentheses denotes the system number. The systems are sorted according to accuracy. Below the horizontal line are results from those literature studies for which the same performance measures were reported (or could be computed). Note that the test sets in these studies are different.

Discussion

We start with a discussion about the performance of the “simple” systems. Assigning all pixels to the most likely class, the background, yields a 73.6% accuracy. Duryea, who calculated the same number on his dataset of 802 images, reported a background percentage of 75.1. Taking the most likely class as a function of location (segmentation independent of the input image) gives an accuracy of 89.8%. The same method was applied by Duryea and Boone and by Vittitoe *et al.* and they reported 87.9% and 88.0%, respectively. The similarity of these results, although different image sets were used, is striking.

Our result using only intensity as feature, which can be seen as using an optimal multiple threshold, yields an accuracy 84.7%. Vittitoe determined a single optimal threshold for his dataset, which scored 80.6%. Using a neural network classifier with the intensities of a 3 by 3 neighborhood (with the images subsampled to 64 by 64 pixels) of each pixel as input vector, he achieved 89.3% correct classification. The fact that our result with intensity as a single feature scores half-way between those results indicates that the performance increase of the 3 by 3 intensity classification versus fixed thresholding is partly due to the fact that a classifier determines multiple

thresholds and partly due to the textural context information obtained from the pixel neighborhood.

The result of the matching system, accuracy 90.7%, is disappointing, since it hardly outperforms the “average lung field” method. In principle, the matching system can be improved and expanded in several ways. Clearly, a different reference image would give different results. One could test a range of reference images and use the one with the best overall performance. If some information about the chest radiograph is known beforehand, such as the age and gender of the subject, one could use a reference image with similar characteristics. Since the reference image should contain features common to all chest radiographs but not those particular to just one, one could take the average image of a range of matched chest images. A more involved alternative would be to do the matching with several reference images and keep the one with the highest similarity measure as the final result. However, pilot experiments with that approach did not lead to much improvement. Varying matching parameters such as the mask size (feature space) and using different similarity

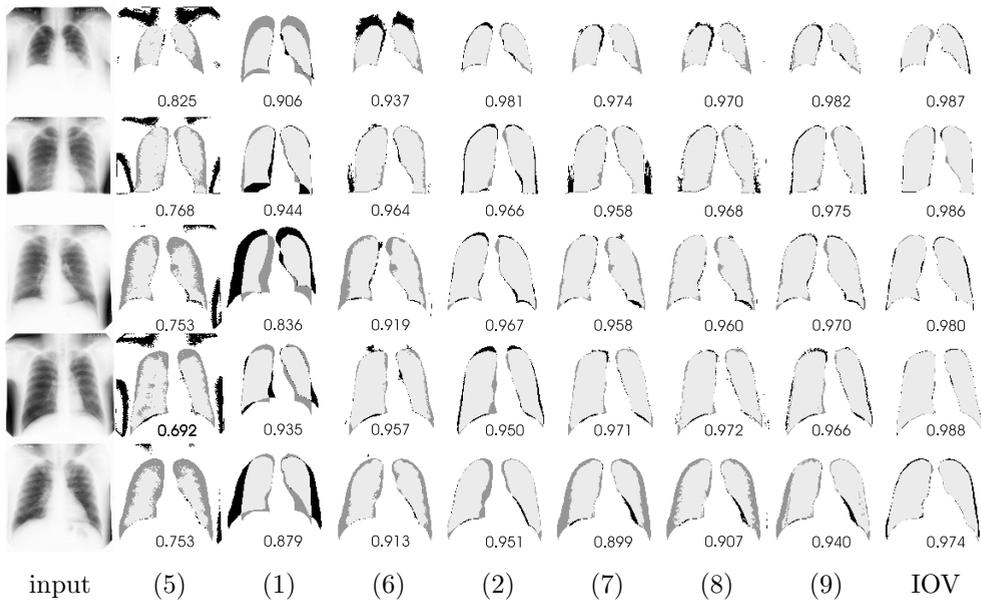


Figure 3.6: Segmentation results for 5 radiographs, shown in the left column, for several systems, indicated in parentheses under each column. The right column shows the inter-observer variability (a comparison of the segmentations of two observers). True negative pixels are shown white, true positive light gray, false positive pixels are dark gray and false negative pixels are shown in black. All 115 test images were sorted based on the accuracy of system 9, the most accurate system and assigned a position, 1 for the highest accuracy, 115 for the lowest accuracy. The top row shows image 1 (highest accuracy), the bottom row shows image 115 (lowest accuracy). The second to fourth row show the images at positions 23, 57, and 80, respectively.

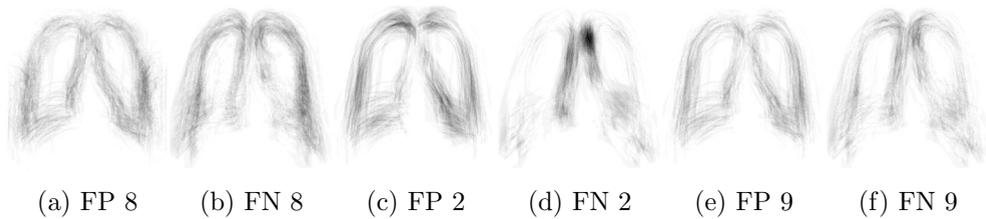


Figure 3.7: False positives and false negatives as a function of location. The maps shown are for systems 8, 2, and 9 respectively. The darker a pixel, the higher the chance that the classification of a pixel at that location is false positive and false negative, respectively. The intensity scaling is the same for each image, so one can compare the performance of each system on these maps directly. Pure black in these images corresponds to incorrect classification of 33.0%.

metrics, such as mutual information and histogram energy, gave similar results in all cases. In principle definite performance improvement could be obtained by allowing elastic transformations. It can be seen easily by inspecting some chest radiographs that even an optimal affine transformation will not result in a perfect match between images.

Now let's consider the inter-observer variability. This score provides a theoretical upper bound. The inter-observer variability is mainly due to the difficulty in assessing the exact borders of mediastinum. Given the performance of "simple" systems and the upper bound set by the inter-observer variability, a segmentation scheme for lung fields in chest radiographs will in practice have an accuracy between 90% and 98%.

It is surprising that a classifier using intensity and the location of pixels as feature (system 6) already gives an accuracy of 93.3%, especially if this is compared with the results of McNitt-Gray [178] who obtained 93.2% using 59 features and spatial information. We do not think that our use of a k NN-classifier instead of a neural network causes the good performance of our pixel classifiers; McNitt-Gray concluded [179] that neural networks only slightly outperformed k NN-classifiers on his data. It may be partly explained by the difference in database. McNitt-Gray used a test and training set of only 17 images each and our much larger database may contain more statistical information. However, it may also indicate that, apart from spatial information, the use of many more local features than (raw) intensity, can hardly improve the discriminating power of pixel classifiers in segmenting lung fields.

There are several arguments in favor of this hypothesis. Our rule-based scheme, and Duryea's rule-based scheme, outperforms McNitt-Gray's 59 feature classifier and Vittioe's pixel classification using Markov random field modeling. In this study, little improvement is obtained by adding entropy as a feature to our classifier; the accuracy of system 8 is only 0.3% higher than that of system 7. And when using 8 features out of his total set of 59, the performance of McNitt-Gray's classifier did not decrease much.

The rule-based system (system 2) is accurate and robust, but there is still room for improvements. As to be expected, adding rules seems the straightforward way to higher accuracy. We distinguish two types of error: *occasional* and *structural*.

The false positive and false negative maps, Figure 3.7c and 3.7d, provide a good way to locate structural errors. It is evident that the detected edge of left lung and mediastinum in the upper lung part is structurally too much to the right. Note that the false positive and false negative maps for systems based on classifiers 3.7(a-b) and 3.7(e-f) do not show such pronounced locations of structural errors. This is an example of a situation where classifiers benefit from their guaranteed optimal performance, contrary to rule-based schemes where the intuition of the designer may fail. Other structural errors are the detection of the outer rib cage, which is detected as a line structure and located on the center of the overlapping ribs. The lung fields end slightly more medial. Because of the coarse scales used to detect the lung edges, their location has shifted. In principle, one could use edge focusing techniques [19] to trace the edges back to a finer scale. In practice, this procedure introduces new scale parameters and does not easily lead to improved results, especially not if the edges are curved or connected.

The most important categories of occasional errors, of which examples are given in Figure 3.8, are

- Complicated patterns of stomach gasses and cardiac edges are common near the lower left lung fields. Therefore the detection of the left diaphragm occasionally fails. This occurred in 18 of the 115 cases. The failure may have profound effects, as shown in Figure 3.8(top-left), or minor consequences (Figure 3.8(top-right)).
- Another common situation is dense lower lung fields, usually due to poor inspiration of the subject. In 9 cases this leads to failure to detect the left diaphragm (Figure 3.8(bottom-left)).
- In two cases, the lung tops were not accurately detected. This was caused by poor visibility of the rib cage. An example is shown in Figure 3.8(bottom-middle).
- Large hila or dense medial lung fields may cause inaccuracies in the detection of the mediastinum edges. Partly this is a structural problem as well. An example is shown in Figure 3.8(bottom-right).

Increasing the robustness of the left diaphragm detection could significantly improve the segmentation method. Indications for failure are differences between the average height of right and left diaphragm and the presence of diaphragm-like structures above/below the detected diaphragm (in case of unexpectedly low/high left diaphragms). Adding rules is the obvious way to extend a rule-based system and this has been noted by Xu [270] who developed tests to classify thoraxes into several categories based on the stomach gasses appearance. Unfortunately, he did not evaluate his scheme in terms of accuracy, sensitivity and specificity, so we cannot compare the performance of his algorithms with ours. Rule-based reasoning is a proper tool to tackle this problem. Vittitoe [250] states that textural features may discriminate between gassy stomach and lung regions. Although this might be true in some cases, in general we expect that local analysis will be unable to make the distinction because *through* the stomach gasses, one discerns lung texture.

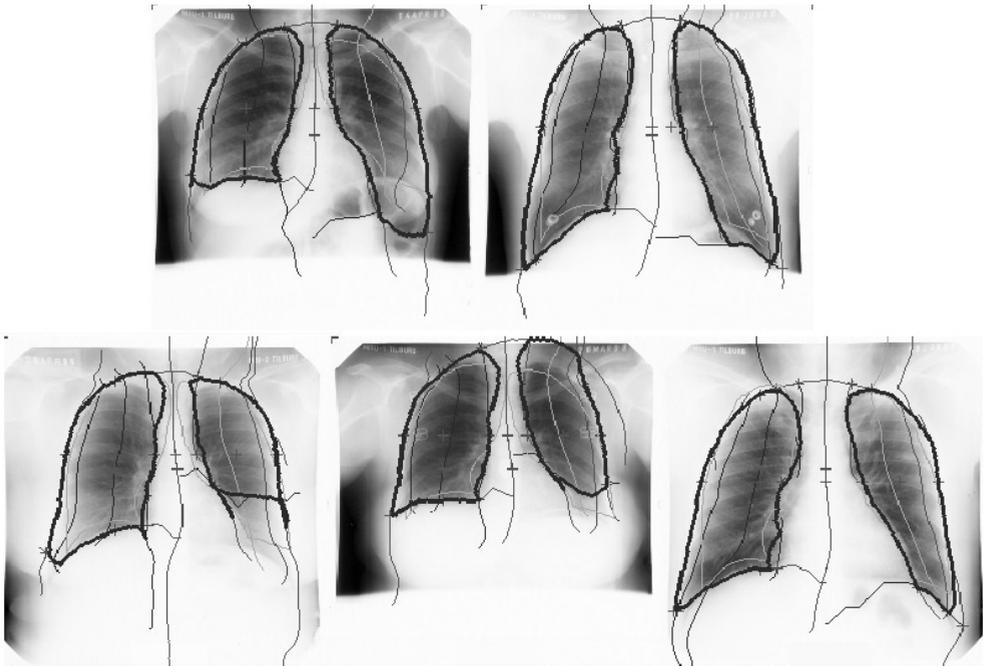


Figure 3.8: Overview of occasional failures of the rule-based segmentation scheme. (Top row) Largely and slightly incorrect detection of left diaphragm due to the presence of stomach gasses (occurred in 18 of 115 cases). (Bottom-left) Incorrect detection of left diaphragm due to a dense low left lung field (2 cases). (Bottom-middle) Incorrect detection of the lung top (9 cases). (Bottom-right) Incorrect detection of the right mediastinum edge due to a pronounced right hilum.

The remaining pixel classifiers (Systems 7 and 8) use the corrected location as feature. Since these locations are calculated based on the results of the rule-based scheme, these systems combine rule-based results with pixel classification. Nevertheless, their performance does not exceed that of the rule-based method. System 8 is comparable to the approach of Tsujii [235], who used a corrected location, an intensity feature and an entropy measure. The fact that he obtains lower accuracy may be due to a different database, but we must also note that his scheme to determine the corrected location is rather simple, compared to our approach. The brightness of the radiographs varies per subject (although the Electrodela chest unit performs some correction to obtain equally bright images), one could attempt to correct for this by scaling the intensity values relative to certain areas that are assumed to be in the lung fields or in the mediastinum. This was done by Tsujii. Another alternative could be to use the entry of a pixel in the cumulative histogram, instead of its intensity as feature. McNitt-Gray investigated the use of this feature. This is equivalent to replacing radiographs with histogram equalized versions. To us this seems rather odd since the amount of dark areas outside the body, varies significantly per image, and affects where pixels within the lung fields end up in the cumulative histogram.

In pilot experiments, we did not note any performance gain using such corrections.

Using corrected location alone is not sufficient to outperform the rule-based method with a classifier. Therefore we attempted to combine the strengths of both methods in a system that reclassified those pixels for which rule-based segmentation and pixel classification with corrected location and intensity gave different results. This (system 9) increased the accuracy from 0.961 to 0.969. This improvement may seem small, but in fact it bridges 35% of the gap between the rule-based scheme (96.1%) and the maximally achievable inter-observer variability of 98.4%, and therefore may be considered a significant improvement. The result can be appreciated in Figure 3.7. Note that the structural errors of the rule-based system (Figure 3.7(d)) have decreased while at the same time the overall performance is better than that of the pixel classifier of system 8 (Figure 3.7(a) versus (3.7(e) and 3.7(b) versus 3.7(f)).

Even more important than the increase in average accuracy, sensitivity and specificity, is the gain in robustness. The standard deviation drops by 30% relative to the rule-based system. Reclassification is a means to correct those cases where the rule-based system makes an occasional gross failure. If we consider the 10% of images that scored worst with the rule-based scheme, the average accuracy is 93.7%. For the corrected scheme this percentage is 95.2%. Note that these are different images, so for the 10 worst cases for the rule-based method, the performance gain is even larger. The worst case of all 115 images using the rule-based system scored 90.1%, for the corrected scheme the worst result was 94.0%. The increase in robustness is also visualized in Figure 3.9, where one can observe that the difference in accuracy between the corrected scheme and schemes 2 and 7 is largest for those images that are relatively difficult to segment.

Note that only pixels classified differently by systems 2 and 7 are eligible for reclassification. Clearly this does not include all incorrectly classified pixels, and thus this system may be improved by using a more elaborate reclassification scheme. Another improvement could be to take into account the fact that lung regions cannot contain holes or ragged edges, which can result from pixel classification. This could be implemented using a final matching step on the result of system 9, just as was done in order to convert the contours detected by the rule-based system into closed lung field contours.

The implementations we used were not aggressively optimized for speed, although the k NN-implementation by Mount and Arya [8] available at <http://www.cs.umd.edu/~mount/ANN> is very efficient. There is ample room for speed improvements. First of all, subsampled versions of the input images can be used, after which for higher resolution images only pixels close to the lung contours need to be classified. For the rule-based scheme, 90% of the computation time is needed to compute the various derivatives of the input image. For most derivatives, these computations are only required for a small part of the image. An implementation that takes advantage of this fact, and possibly uses fast (approximate) implementations of derivatives, will be several orders of magnitude faster. Furthermore, the classification times can be reduced by allowing *approximate* nearest neighbors [8] or even using look-up tables. The hybrid scheme uses both the rule-based segmentation and the pixel classifier from system 7. These two calculations are independent and could be implemented in parallel on a dual processor PC. Taken this together, the segmentation of a single

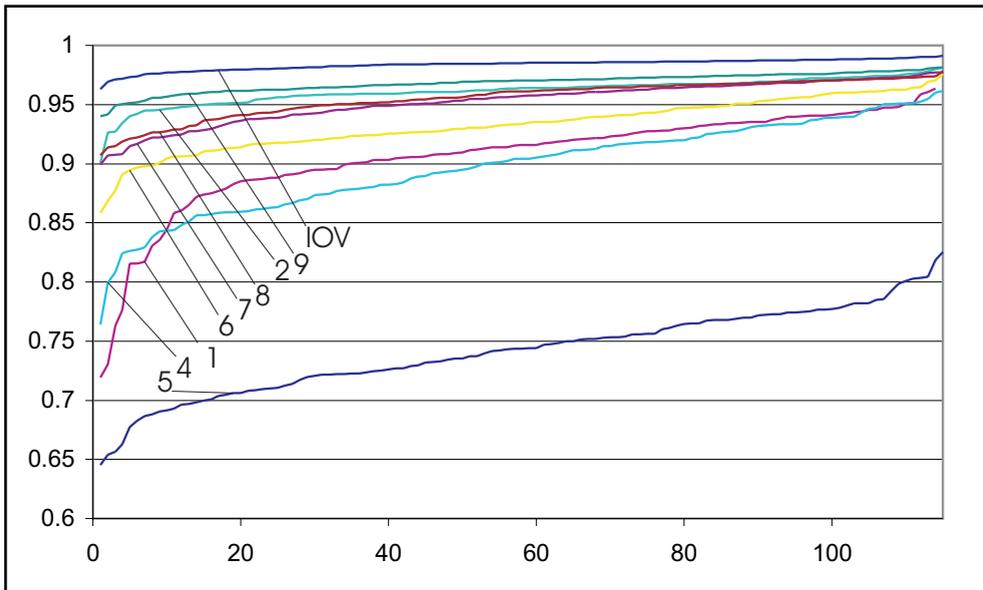


Figure 3.9: Illustration of the differences in performance of each system, by plotting the accuracy of all images sorted in ascending order. Each line represents a system, as indicated in the graph. IOV is the line for the inter-observer variability. To the left of the graph, where images that are hard to segment are plotted, the differences between the various systems are most clear.

input image, even at a full resolution used in clinical practice, can be done within a second. This allows for application in clinical and mass chest screening practice.

Conclusions

One of the goals of this study was to compare the performance of various approaches to the segmentation of lung fields in chest radiographs. Because of the detailed comparisons between the 9 systems considered here, we believe we may have shed light on which parts and elements of the segmentation schemes proposed in this work and previous studies are actually important for their performance in practice. It turns out that both pixel classification with simple features and rule-based schemes focused on contour detection can be used to obtain good results. All in all, rule-based lung contour detection seems to perform slightly better than pixel classification. This indicates that the use of global information, to which pixel classification approaches are not directly suited, is important for lung field segmentation. The rule-based scheme we have developed in this work is robust and accurate, but can still be improved in several ways, notably by adding rules to handle exceptional situations. We have shown that both methods are complementary, because they can be combined into a hybrid scheme that is both more robust and more accurate. This accuracy of this

scheme approaches the theoretical limit of inter-observer variability and seems to outperform algorithms published so far, although this remains to be proven by an extensive evaluation of several algorithms on one, common database.

Chapter 4

Applications of Locally Orderless Images

In a recent work [141], Koenderink and Van Doorn consider a family of three intertwined scale-spaces coined the *locally orderless image* (LOI). The LOI represents the image, observed at inner scale σ , as a local histogram with bin-width β , at each location, with a Gaussian-shape region of interest of extent α . LOIs form a natural and elegant extension of scale-space theory, show causal consistency and enable the smooth transition between pixels, histograms and isophotes. The aim of this work is to demonstrate the wide applicability and versatility of LOIs. We present applications for a range of image processing tasks, including new non-linear diffusion schemes, adaptive histogram equalization and variations, several methods for noise and scratch removal, texture rendering, and texture segmentation.

HISTOGRAMS ARE UBIQUITOUS in image processing. They embody the notion that for many tasks the distribution of intensity values within a region of interest contains the required information, with the spatial order of the pixels being immaterial. One can argue that even at a single location (which is a physical impossibility since any measurement is of non-zero scale) the intensity has an uncertainty, and should therefore be described by a probability distribution: physical plausibility requires non-zero imprecision. This led Griffin [89] to propose a *scale-imprecision space* with spatial scale parameter σ and an intensity, or tonal scale β , which can be identified as the familiar *bin-width* of histograms.

Koenderink and Van Doorn [141] extended this concept to *locally orderless images* (LOIs), an image representation with three scale parameters in which there is no local but only a global topology defined. LOIs are *local histograms*, constructed according to scale-space principles, *viz.* without violating the causality principle [136]. As such, one can apply to LOIs the whole machinery of techniques that has been developed in the context of scale-space research, such as the N-jet for local image structure description [137, 70, 94], the construction of invariants and (oriented) filter families [138, 139, 140, 71], non-linear diffusion schemes [195, 93], scale selection methods [160,

161], etc.

The aim of this chapter is to demonstrate that LOIs are a versatile and flexible framework for image processing applications; it may be viewed upon as a broad feasibility study. We do not give thorough evaluations for each of the applications presented. Obviously, local histograms are in common use in many areas, and the notion to consider histograms at different scales (soft binning) is not new either. Yet we believe that the use of a consistent mathematical framework in which all scale parameters are made explicit can aid in the design of effective algorithms by reusing existing scale-space concepts. Additional insight may be gained by taking into account the behavior of LOIs over scale.

Locally orderless images

We first briefly review locally orderless images [141] and comment on some interesting attributes. A convenient introduction is to consider the scale parameters involved in the calculation of a histogram:

- the inner scale σ with which the image is observed;
- the outer scale, or extent, or scope α that parameterizes the size of the field of view over which the histogram is calculated;
- the scale at which the histogram is observed, tonal scale, or bin-width β .

First of all, note that these scale parameters are distinct. Increasing σ , i.e. blurring the image, has a different effect on the histogram than increasing β , i.e. blurring the histogram itself. Consider a black-and-white image. The histogram contains two peaks and blurring the histogram will change this eventually to a uniform distribution. Blurring the image, on the other hand, will eventually create a uniform image, whose histogram is a delta peak.

The locally orderless images $H(\mathbf{x}, i; \sigma, \alpha, \beta)$ are defined as the family of histograms, i.e. a function of the intensity i , with bin-width β of the image observed at scale σ calculated over a field of view centered around \mathbf{x} with an extent α . The unique way to decrease resolution without creating spurious resolution is by convolution with Gaussian kernels [136, 261]. Therefore Gaussian kernels are used for both σ , α and β .

We summarize this with a recipe for calculating LOIs:

1. Choose an inner scale σ and blur the image $L(\mathbf{x}; \sigma)$ using the diffusion

$$\Delta_{(\mathbf{x})}L(\mathbf{x}; \sigma) = \frac{\partial L(\mathbf{x}; \sigma)}{\partial \sigma^2}. \quad (4.1)$$

with the boundary condition that $L(\mathbf{x}; 0)$ is the input image.

2. Choose a number of (equally spaced) bins of intensity levels i and calculate the “soft isophote images”, representing the “stuff” in each bin through the Gaussian gray-scale transformation

$$R(\mathbf{x}, i; \sigma, \beta) = \exp\left(-\frac{(L(\mathbf{x}; \sigma) - i)^2}{2\beta^2}\right) \quad (4.2)$$

3. Choose a scope α for a Gaussian aperture, with unit amplitude

$$A(\mathbf{x}; \alpha) = \exp\left(-\frac{\mathbf{x}^2}{2\alpha^2}\right) \quad (4.3)$$

and compute the locally orderless image through convolution

$$H(\mathbf{x}, i; \sigma, \alpha, \beta) = \frac{A(\mathbf{x}; \alpha)}{2\pi\alpha^2} \otimes R(\mathbf{x}, i; \sigma, \beta). \quad (4.4)$$

Note that $R(\mathbf{x}, i; \sigma, \beta)$ and $H(\mathbf{x}, i; \sigma, \alpha, \beta)$, are stacks of isophote images, and therefore have a dimensionality 1 higher than that of the input image.

4. The bin-width β can be increased by keeping \mathbf{x} , σ , and α fixed and blurring along the intensity direction:

$$\frac{\partial^2 H(\mathbf{x}, i; \sigma, \alpha, \beta)}{\partial i^2} = \frac{\partial H(\mathbf{x}, i; \sigma, \alpha, \beta)}{\partial \frac{\beta^2}{2}}, \quad (4.5)$$

5. The extent α can be increased by keeping i , σ , and β fixed and blurring each “intensity slice”:

$$\Delta_{(\mathbf{x})} H(\mathbf{x}, i; \sigma, \alpha, \beta) = \frac{\partial H(\mathbf{x}, i; \sigma, \alpha, \beta)}{\partial \frac{\alpha^2}{2}}, \quad (4.6)$$

For an illustration of the construction of an LOI, see Figure 5.1. The term *locally orderless* image refers to the fact that we have at our disposal at each location the probability distribution, which is a mere orderless set; the spatial structure within the field of view α centered at \mathbf{x} has been completely disregarded. This is the key point: with each spatial location we do not longer associate a (scalar) intensity, but a probability distribution, parameterized by σ , α , β . Since a distribution contains more information than the intensity sec, we may expect to be able to use this information in various image processing tasks.

The LOI contains several conventional concepts, such as the original image and its scale-space $L(\mathbf{x}; \sigma)$ that can be recovered from the LOI by integrating $iH(\mathbf{x}, i; \sigma, \alpha, \beta)$

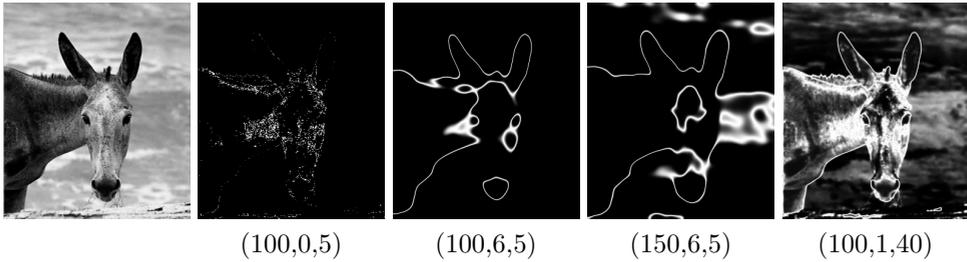


Figure 4.1: Examples of soft isophote images. The input image on the left has size 310×357 pixels and intensity values are distributed in the range $[0,255]$. The triple of numbers below each image indicates the intensity value i , inner scale σ and tonal scale β , respectively. Note that, because $\beta > 0$ the isophotes are indeed *soft*: at some positions they fade gradually from white to black, at other positions this transition is very sharp.

over i . The image histogram in the conventional sense is obtained by letting $\alpha \rightarrow \infty$. The construction also includes families of isophote images, which for $\beta > 0$ are named *soft isophote images* by Koenderink (see Figure 4.1). And maybe even more important, by tuning the scale parameters the LOI can fill intermediate stages between the image, its histogram and its isophotes. This can be useful in practice. Finally, we note that the framework generalizes trivially to n D images or color images, if a color metric is selected.

Median and maximum mode evolution

If we replace the histogram at each location with its mean, we obtain the input image $L(\mathbf{x}; \sigma)$ blurred with a kernel with width α . This holds independently of β , since blurring a histogram does not alter its mean. If, however, we replace the histogram with its median or its maximum mode (the intensity with the highest probability), we obtain a diffusion with scale parameter α that is reminiscent of some non-linear diffusion schemes (see Figure 4.2). In fact it has been shown by Guichard and Morel [91] that mean curvature flow corresponds to iterated median filtering, but the relation between other histogram operations and non-linear diffusion schemes is less clear. The tonal scale β works as a tuning parameter that determines the amount of non-linearity. Griffin [89] proved that for infinite imprecision ($\beta \rightarrow \infty$) median filtering corresponds to linear filtering. With only a few soft isophote level images in the LOI, maximum mode diffusion also performs some sort of quantizing, and one obtains piecewise homogeneous patches with user-selectable values. This can be useful, *e.g.* in coding for data compression and knowledge-driven enhancements.

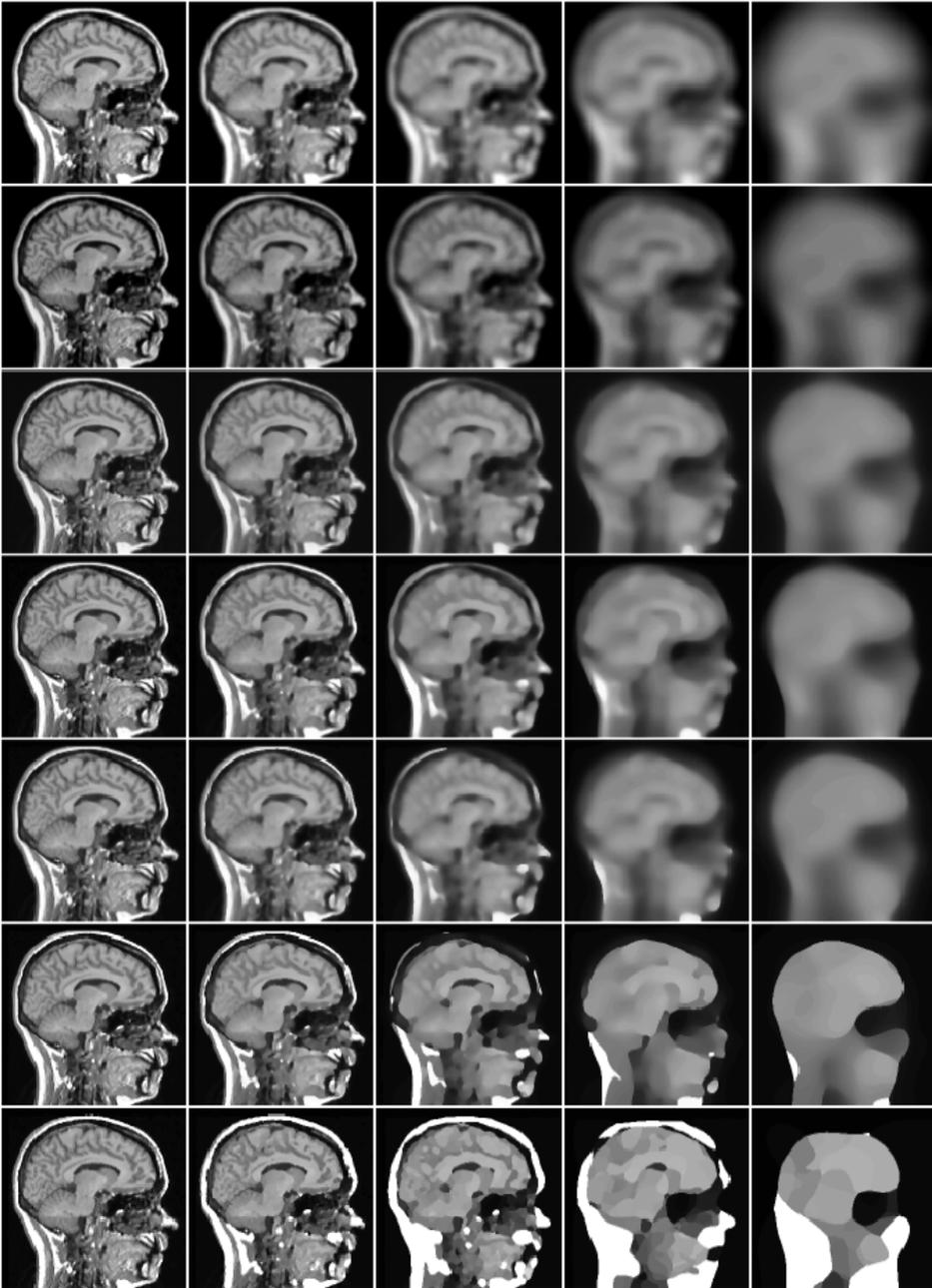


Figure 4.2: Input image is a sagittal 256×256 MR brain image with intensities in the range $[0, 255]$. Each row shows a diffusion of the LOI with $\sigma = 0$ for $\alpha = 1, 2, 4, 8, 16$. Top row: mean diffusion, equivalent to linear scale space. Row 2 to 4: median diffusion for $\beta = 64, 16, 1$. Row 5 to 7: maximum mode diffusion for $\beta = 64, 16, 1$.



Figure 4.3: Left: text hidden in a sinusoidal background, dimensions 230×111 , intensities in the range $[0, 1]$. Middle: bi-modal locations in an LOI of $\sigma = 0$, $\beta = 0.15$ and $\alpha = 1.5$. Right: bi-modal locations have been replaced with the high mode. Text is removed and the background restored.

Switching modes in bi-modal histograms

Instead of replacing each pixel with a feature of its local histogram, such as the median or the maximum mode, we can perform more sophisticated processing if we take the structure of the local histograms into account. If this histogram is bi-modal, this indicates the presence of multiple “objects” in the vicinity of that location.

Consider locations with bi-modal histograms (histograms with two maxima, with a single minimum in between). We let pixels at such locations “switch” to a desired mode. That is, if they are in the high mode (their value is higher than that of the minimum in between the two modes), we replace their value with the low mode (or vice versa, depending on the desired effect). The idea behind this is that a bright/dark object is replaced with the most likely value that the darker/brighter object in its vicinity has, namely the low/high maximum mode. Note that this a two-step process: the detection of bi-modal locations is a segmentation step, and replacing pixels fills in a value from its vicinity, thus using statistical information, and taking into account only those pixels that belong to the object to be filled in.

This scheme allows for a scale-selection procedure in the following way. For fixed σ, β, α , there may be locations with more than two modes in their local distribution. This indicates that it is worthwhile to decrease α , focusing on a smaller neighborhood, until just two modes remain. Thus we use a locally adaptive α , ensuring that the replaced pixel value comes from information from locations “as close as possible” to the pixel to be replaced.

We have applied this scheme successfully for the removal of text on a complicated background (Figure 4.3), the detection of dense objects in chest radiographs, and noise removal. Figure 4.4 shows how shot noise can be detected and replaced with a probable value, obtained from the local histogram. The restoration is near perfect. We also applied the mode switching technique to old movie sequences with severe deteriorations. Most artifacts could be removed automatically.

Histogram transformations

Any generic histogram can be transformed into any other histogram by a nonlinear, monotonic gray-level transformation. To see this, consider an input histogram $h_1(i)$

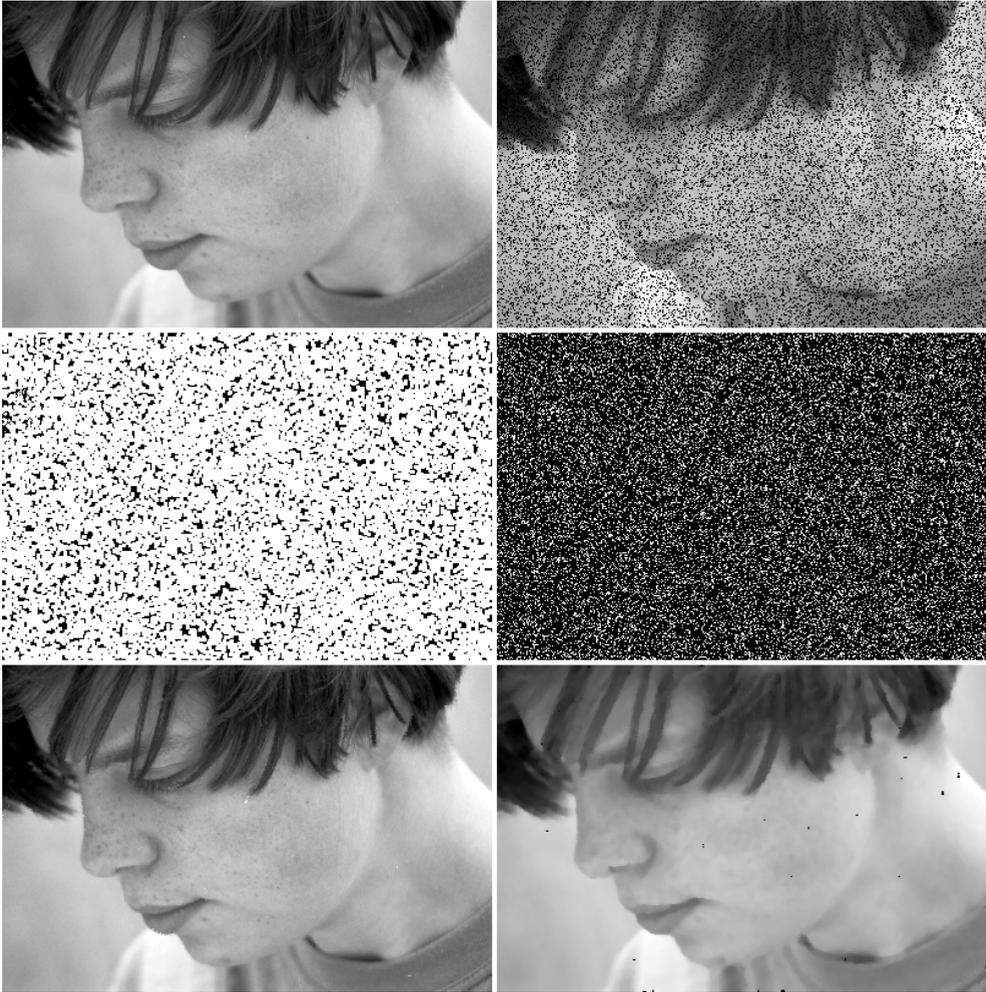


Figure 4.4: Top-left: original image, 400×267 pixels, intensities scaled to $[0, 1]$. Top-right: original image with 20% shot noise. This is the input image for the restoration procedure. An LOI with $\sigma = 0$, $\beta = 0.04$, $\alpha = 0.4$ was computed. Middle-left: binary image with locations in top-right image with bi-modal histograms shown in white. Middle-right: binary image with locations in top-right image with bi-modal histograms and pixels in the low mode shown in white. These locations are replaced. Bottom-left: restoration using mode-switching for bi-modal locations gives excellent results. Bottom-right: restoration using a 5×5 median filter removes most but not all shot noise, and blurs the image.

and its cumulative histogram $H_1(i) = \int_{-\infty}^i h_1(i')$ and the desired output histogram $h_2(i)$ and $H_2(i)$. If we replace every i with the i' for which $H_1(i) = H_2(i')$ we have transformed the cumulative histogram H_1 into H_2 and thus also h_1 into h_2 . Since cumulative histograms are monotonically non-decreasing, the mapping is monotonically

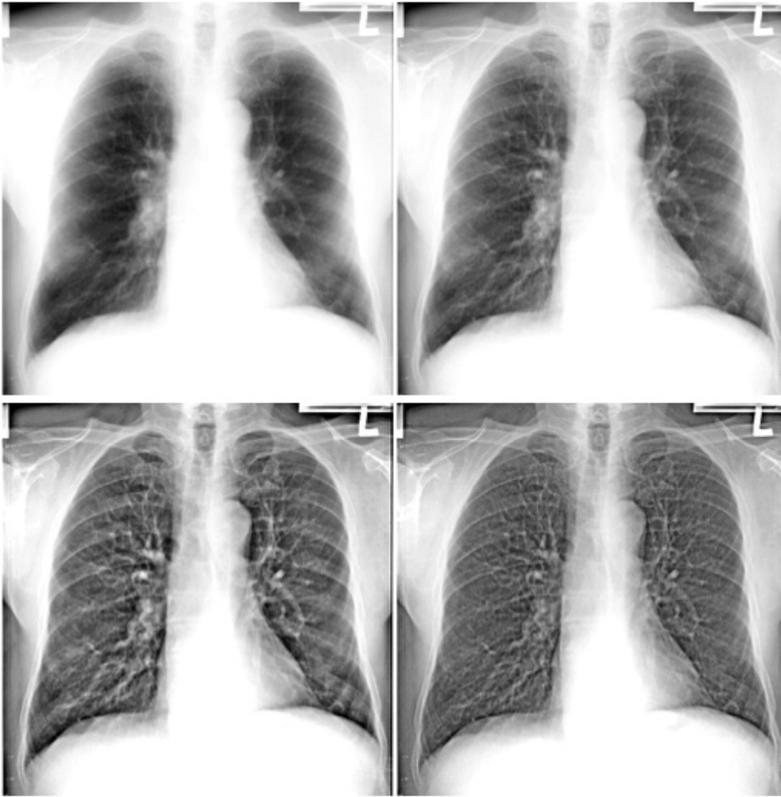


Figure 4.5: A normal PA chest radiograph of 512 by 512 pixels with intensities in the range $[0,1]$. Top-left: original image, in which details in lung regions and mediastinum are not well visible due to the large dynamic range of gray levels. Top-right: adaptive histogram equalization (AHE) based on the LOI with $\sigma = 0$, $\alpha = 8$; $\beta = 0.4$. Bottom-left: AHE with $\sigma = 0$, $\alpha = 8$; $\beta = 0.2$. Bottom-right: AHE with $\sigma = 0$, $\alpha = 4$; $\beta = 0.2$.

non-decreasing as well.

Although in practice this can only be achieved to limited accuracy due to discretization of the intensity domain, histogram manipulation is a very useful technique if one has an idea of what the output histogram as a function of the input histogram should be.

Histogram equalization is an example. The idea is that when displaying an image with a uniform histogram (within a certain range), all available gray levels or colors will be used in equal amounts and thus “perceptual contrast” is maximal. The idea to use local histograms (that is, selecting a proper α for the LOI) for equalization, to obtain optimal contrast over each region in the image stems from the 1970s [108] and is called *adaptive histogram equalization* (AHE). Several variations on this theme have been proposed for pre-processing of medical images, notably chest radiographs [35].

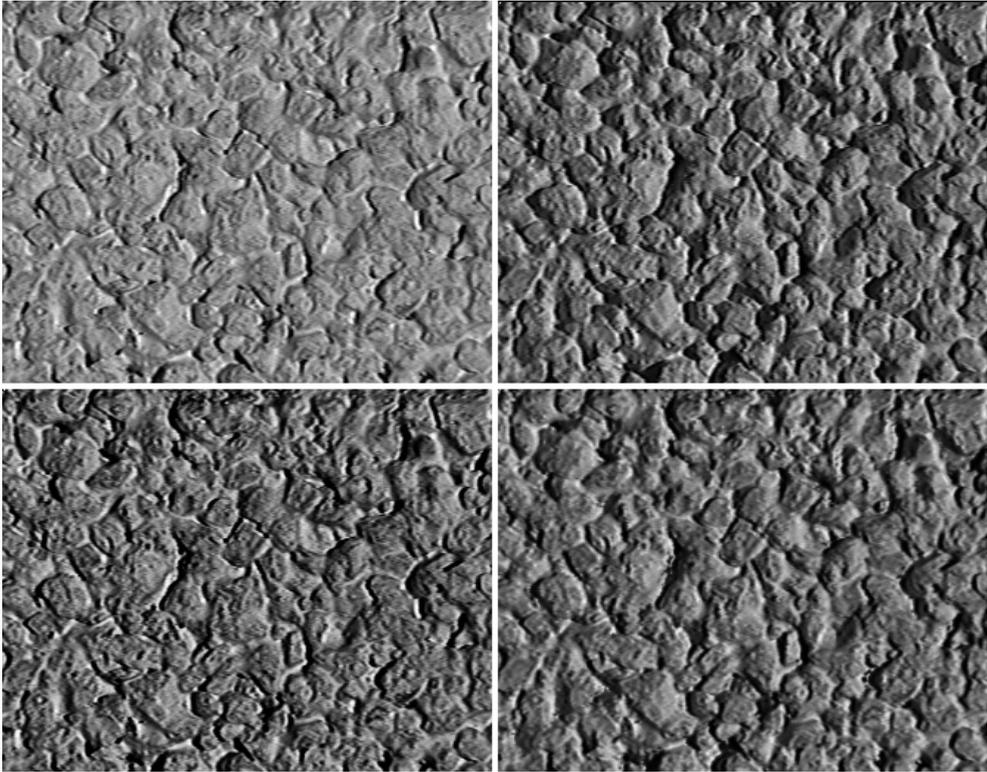


Figure 4.6: Top-left is an image (332 by 259 pixels) of rough concrete viewed frontally, illuminated from 22° . Top-right: the same material illuminated from 45° . Bottom-left shows the top-left image with its histogram mapped to the top-right image to approximate the change in texture. Bottom-right shows the result of local histogram transformation, with $\alpha = 2$. The approximation is especially improved in areas that show up white in the images on the left. These areas are often partly shadowed with illumination from 45° , and using a local histogram may correctly “predict” such transitions.

However, it was noted that these operations blow up noise in homogeneous regions. Pizer et al. [199] proposed to *clip* histograms, *viz.* for each bin with more pixels than a certain threshold, truncate the number of pixels and redistribute these uniformly over all other bins. It can be seen that this ad hoc technique amounts to the same effect as increasing β in the LOI; notably, for $\beta \rightarrow \infty$, AHE has no effect. Thus we see that the two scale parameters α and β determine the size of structures that are enhanced and the amount of enhancement, respectively. Figure 4.5 shows a practical example of such a continuously tuned AHE for a medical modality (thorax X-ray) with a wide latitude of intensities.

An alternative to histogram equalization is to increase the standard deviation of the histogram by a constant factor, which can be done by a linear gray level transformation. Again, the LOI provides us with an elegant framework in which the

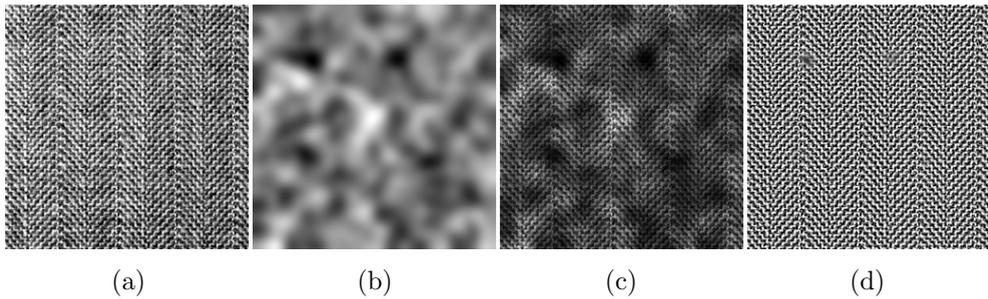


Figure 4.7: (a) A texture from the Brodatz set [26], resolution 256^2 , intensities in the range $[0, 1]$. (b) Blurred Gaussian noise, scaled to range from $[0, 1]$. (c) Multiplication of (a) and (b). (d) Reconstruction of (a) from (c) from the LOI with $\sigma = 0, \beta = 0.1, \alpha = 2$ and computing for each point a mapping to the local histogram at the randomly chosen location $(80, 80)$.

scale parameters that determine the results of such operations are made explicit.

Another application of histogram transformation is to approximate changes in texture due to different viewing and illumination directions [81]. In general, the textural appearance of many common real-world materials is a complex function of the light field and viewing position. In computer graphics it is common practice, however, to only apply a projective transformation to a texture patch in order to account for a change in viewing direction and to adjust the mean brightness using a bi-directional reflection distribution function (BRDF), often assumed to be simply Lambertian. In [81] it is shown that this gives poor results for many materials, and that histogram transformations often produce far more realistic results. A logical next step is to consider *local* histogram transformations. An example is shown in Figure 4.6, using a texture of rough concrete taken from the CURET database [53]. Instead of using one mapping function for all pixel intensities, the mapping is now based on the pixel intensity and the intensities in its surroundings. Simple physical considerations make clear that this approach does make sense: bright pixels which have dark pixels due to shadowing in their neighborhood are more likely to become shadowed for more oblique illumination than those that are in the center of a bright region.

Finally, histogram transformations can be applied to restore images that have been corrupted by some noise process, but for which the local histogram properties are known or can be estimated from the corrupted image. Although this may sound contrived, we believe such cases are encountered frequently in practice. Many image acquisition systems contain artifacts that are hard to correct with calibration schemes. One example in medical image processing is the inhomogeneity of the magnetic field of an MR scanner or of the sensitivity of MR surface coils, leading to low frequency gradients over the image, for which retrospective correction schemes have been proposed [156, 157]. A generated example is shown in Figure 4.7 where we multiplied a texture image with Gaussian noise. We make two assumptions: 1) the noise does not contain high frequency components and thus the structure of the local histogram will

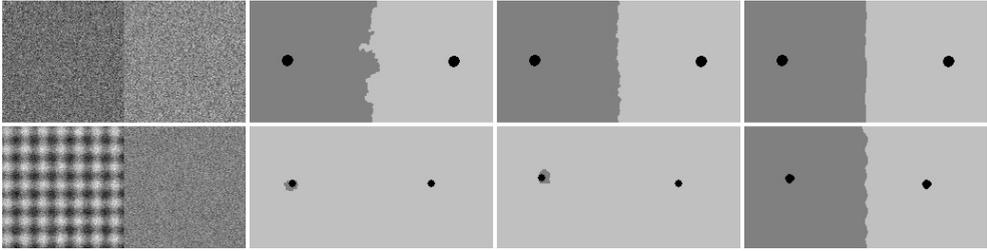


Figure 4.8: Top row, from left to right: a 256×128 test image composed of two homogeneous regions with intensity 0 and 1 and Gaussian noise with zero mean and unit variance. An LOI with $\sigma = 0$ and $\beta = 0.2$ and $\alpha = 0, 1, 4$, respectively is used for seeded region growing from the two seeds shown in black. Since the mean of the two regions is different, regular seeded region growing ($\alpha = 0$) works fairly well, although the segmentation improves with increasing α . Bottom row: same procedure for a partly textured image; the left half was filled with $\sin(x/3) + \sin(y/3)$, the right half was set to zero, and Gaussian noise with zero mean and $\sigma = 0.5$ was added. Regular seeded region growing now fails, but if α is large enough, the segmentation is correct.

not be affected for a small scope α ; 2) the original texture image has small textons, so that over a small scope, the histograms are equal everywhere. By randomly choosing a point in the corrupted image and computing the mapping that transforms each local histogram to the local histogram at that particular location we obtain the restored image in Figure 4.7(d). The texture of the restored image is more uniform than in the original, indicating that assumption 2) is not completely valid. Furthermore, in areas where the noise factor with which the texture was multiplied was close to 0, all information is destroyed, and restoration fails. Note that we did not make any assumption about the noise process. In this particular case, dividing Figure 4.7 with a median filtered version of it, would give good restoration as well, but this assumes corruption by multiplicative noise. The LOI method works for additive noise, or other kinds of noise as well.

Texture segmentation based on local histograms

Many “general” (semi-)automatic segmentation schemes are based on the notion that points in spatial proximity with similar intensity values are likely to belong to the same object. Such methods run into problems with textured areas, because the intensity values may show wild local variations. A solution is to *locally* compute texture features and replace pixel values with these features, assuming that pixels that belong to the same texture region will now have a similar value. The framework of LOIs is ideally suited for the computation of such local features. The idea of using texture information sampled from a local aperture is not new. Shi and Malik [220] have applied their normalized cut segmentation scheme to texture segmentation in this way, using local histograms and the correlation between them as a metric. Gårding and Lindeberg [76] used an *integration scale*, comparable to α , to denote the aperture

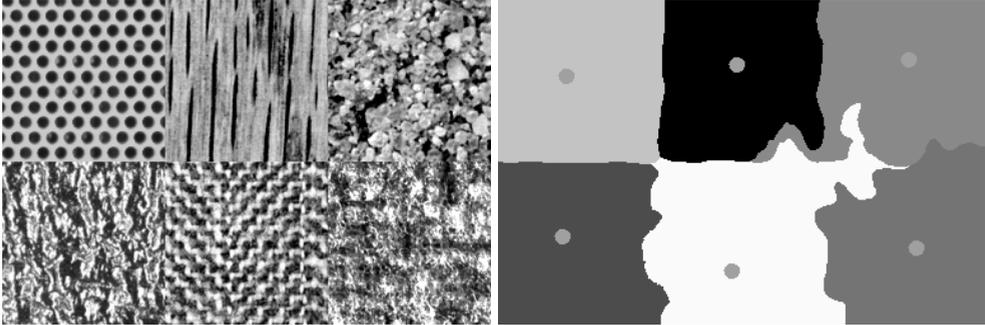


Figure 4.9: Left: a test image composed of 6 texture patches of pixel size 128×128 each. Intensity values in each patch are normalized to zero mean and unit variance. Right: the result of segmentation with seeded region growing based on an LOI with $\sigma = 0$, $\beta = 0.2$ and $\alpha = 8$. The circles are the seeds.

over which texture information is sampled.

Here we present an adapted version of a semi-automatic segmentation technique called seeded region growing (SRG), developed by Adams and Bischof [2], that is popular in medical image processing. An attractive property of our adaptation of this scheme is that it reduces to a scheme very similar to the original SRG scheme for $\alpha \rightarrow 0$. This is directly due to the fact that LOIs *contain* the original image.

SRG segments an image starting from (user-selected or in some way automatically determined) seeds. The algorithm maintains a list of all pixels that are connected to one of the seeds, sorted according to the distance of the pixel to the seed. This metric is originally defined as the intensity of the pixel minus the mean intensity of the seed region squared. The pixel at the top of the list is added to the seed-region it is connected to, the mean intensity of this region is updated and the neighbors of the added pixel are added to the list. This procedure is repeated, usually until all pixels are assigned to a region. The effect is that seeds in a homogeneous region grow quickly; once an edge is reached, the growing process grinds to a halt.

Instead of comparing pixel intensity values with the mean intensity of a region, we compare the local histograms of a pixel and a region. This requires the notion of a distance between two distributions and there are several ways in which this can be implemented. We choose to subtract the histograms and take the sum of the absolute values of what is left in the bins. For $\alpha \rightarrow 0$ this reduces to a scheme similar to the original scheme, except that one considers for the region the global mode instead of the mean of the histogram (most likely pixel value instead of the mean pixel value). Figures 4.8 to 4.10 illustrate the use of seeded region growing based on local histograms.

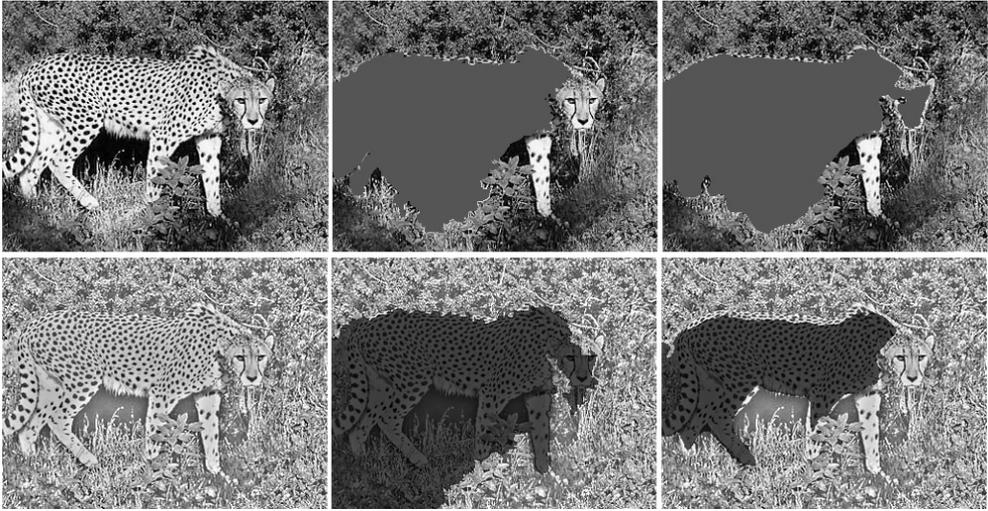


Figure 4.10: Top row, left: wildlife scene with leopard, size 329×253 pixels, intensities scaled between $[0, 1]$; Bottom row, left: a locally ($\sigma = 8$) normalized version of the input image; Middle and right: segmentation by SRG based upon an LOI with $\sigma = 0$, $\beta = 0.05$ and $\alpha = 0, 4$, respectively. Note how well the textured area is segmented in the lower right image.

Concluding remarks

In the applications presented, we have used many aspects of LOIs. In some cases, they are a natural extension ($\alpha > 0$) of techniques that usually use pixels ($\alpha = 0$), *e.g.* seeded region growing. In other cases, they extend techniques that use “conventional” histograms ($\alpha \rightarrow \infty$) with an extra parameter, *e.g.* histogram transformation techniques. Other applications exploit the behavior of LOIs over scale to obtain non-linear diffusions and to remove noise. We conclude that LOIs are image representations of great practical value.

Chapter 5

Texture Classification from Locally Orderless Images

A locally orderless image is a family of three intertwined scale-spaces that describes local histograms. We generalize locally orderless images by considering local histograms of a collection of filtered versions of the image, and by extending them to joint probability distributions. These constructions can be used to derive texture features and are shown to be a more general description of two established texture classification methods, *viz.* filter bank methods and co-occurrence matrices. Because all scale parameters are stated explicitly in this formulation, various multi-resolution feature sets can be extracted in a systematic way. Each multi-resolution approach improves the performance of texture classification schemes, the best result being obtained if a multi-resolution approach for all scale parameters is used. This is demonstrated in experiments on a large data set of 1152 images from 72 texture classes.

HISTOGRAMS or, more generally, joint probability distributions feature prominently in methods for texture feature extraction. The fact that three distinct scale parameters are involved in the calculation of a histogram has practical consequences for these methods. These three scale parameters are the scale at which the image is observed (inner scale σ), the scale of the histogram, which has the unit of the image intensity (tonal scale or bin-width β) and the spatial extent over which the histogram is computed (extent or integration scale α). As a result, histograms are complex objects consisting of three intertwined scale spaces. These objects are called locally orderless images (LOIs) by Koenderink and Van Doorn who have discussed their properties in a recent work [141].

In this paper, we propose two extensions of the LOI framework. First, we consider histograms of a set of images obtained by applying filters to the image. Second, we consider joint probability distributions of locations in the set of filtered images with given spatial relations. It will turn out that these entities describe textural properties of the image in a way similar to filter bank methods and co-occurrence matrices,

respectively, but with all the relevant scale parameters stated explicitly.

In the next section we review related literature on texture classification. Then we present our extensions to LOIs, discuss their properties and show how they can be used to derive texture features. Subsequently, experiments are presented in which a large data set of images of natural and man made textures is classified by ten sets of features. Finally, the results are discussed and conclusions drawn.

Previous work on texture classification

Texture analysis methods have been reviewed in several papers [100,87,204,119]. This section focuses on filter bank methods and co-occurrence matrices.

Broadly speaking, filter bank methods use input images to produce derived images and use statistics of these derived (filtered) images to obtain texture features. A large variety of filter banks have been proposed: differences of offset Gaussians [172]; Gabor filters [24,240], sometimes complemented with the Laplacian of Gaussian filters [280]; transform filters such as the Laws masks [150] and well known transforms such as the Hadamard filters, the discrete sine and cosine transforms, which have been shown to resemble the optimal Karhunen-Loeve transform [241,243,197]; common filters from image processing such as gradient, Laplacian, Sobel operators [68]; mathematical morphology operations and other non-linear filters [104]; Fourier transforms [162] and wavelets [146]. (The given references are by no means complete, and focus on methodological papers.) Motivations for filter banks are usually their properties – in the sense of being complete or orthonormal –, their performance in practice, or evidence of their existence in the human visual system [30].

Deriving features from co-occurrence matrices appears to be the most widely applied method for texture analysis. co-occurrence matrices were first proposed by Haralick [101] and have been the subject of many studies [281,42,43,247,191]. Joint histograms of two locations in the image with pre-described spatial relations are constructed, and usually features are extracted from these matrices.

Filter bank methods and co-occurrence matrices are related in various ways. Unser has shown [242] that co-occurrence matrices can be approximated by combinations of sum and difference histograms, and that most features commonly extracted from co-occurrence matrices can be computed exactly from these sum and difference histograms. Histograms of the sum of two pixel locations are related to the histograms of blurred images; histograms of the difference of two pixels in a given direction are related to histograms of the first derivative of the image in that direction. This shows that there exists a direct relation between the co-occurrence approach and the filter bank approach if the filter bank contains zero-th order and steered first order derivative filters.

Obviously, one can construct co-occurrence matrices from filtered images [54] and thus combine both approaches. A natural generalization is to consider *(i)* a set of filtered images, *(ii)* a number of locations with given spatial relationships, *(iii)* a region of interest, and to extract features from joint probability distributions of intensity values at these locations in the region of interest. This is the generalization that we will propose in this work, with all scale parameters readily identified. In the

case of a single location, this corresponds to filter bank methods. In the case of two locations and the original image as input (a filter bank with just the identity filter) this corresponds to classical co-occurrence matrices. When more than two locations are taken into account, we obtain entities that are similar to those studied by Oja and Valkealahti [192, 244].

Markov Random Field models are another widely used method for texture analysis [52]. Interestingly, Wu [268] has shown that with such a general formulation of filter bank methods and co-occurrence matrices, this approach is equivalent to a special case of an MRF model [280]. This means that it is possible to *generate* textures given the statistics of filter bank responses.

Theory

Locally orderless images

We start by reviewing *locally orderless images* [141] which provide a formal definition of a histogram, consistent with scale-space requirements [136, 160, 69]. (See also Chapter 4 of this thesis.)

Consider the image $L(\mathbf{x}; \sigma)$ that can be observed at any scale σ . The following non-linear gray-scale transformation is applied to $L(\mathbf{x}; \sigma)$:

$$R(\mathbf{x}, i; \sigma, \beta) = \exp\left(-\frac{(L(\mathbf{x}; \sigma) - i)^2}{2\beta^2}\right), \quad (5.1)$$

where i denotes the intensity value. β is a scale parameter in the intensity domain. If we replace $L(\mathbf{x}; \sigma)$ with $R(\mathbf{x}, i; \sigma, \beta)$, we replace pixel values with a probability distribution, with Gaussian shape, centered at the pixel value i with standard deviation β . Thus β can be identified as the scale in the tonal domain, or the *imprecision* of the intensity value [89], or the *bin-width* of a histogram with Gaussian bins.

For fixed i , $R(\mathbf{x}, i; \sigma, \beta)$ is a *soft isophote image*, that indicates the probability that a location in the original image has intensity value i (see also Figure 4.1). The softness of the isophote is regulated by β : For $\beta = 0$ we have a regular isophote image that has either value 1, in which case the location is on the isophote, or 0, in which case the location is not. For a fixed location \mathbf{x} , $R(\mathbf{x}, i; \sigma, \beta)$ is a histogram of the image $L(\mathbf{x}; \sigma)$ for a region of zero size, containing only the pixel at \mathbf{x} . We can increase the extent over which the histogram is calculated by convolving R with a Gaussian aperture, normalized at maximum amplitude 1 and extent α ,

$$A(\mathbf{x}; \alpha) = \exp\left(-\frac{\mathbf{x}^2}{2\alpha^2}\right) \quad (5.2)$$

to obtain the locally orderless image

$$H(\mathbf{x}, i; \sigma, \beta, \alpha) = \frac{A(\mathbf{x}; \alpha)}{(\sqrt{2\pi}\alpha)^D} \otimes R(\mathbf{x}, i; \sigma, \beta), \quad (5.3)$$

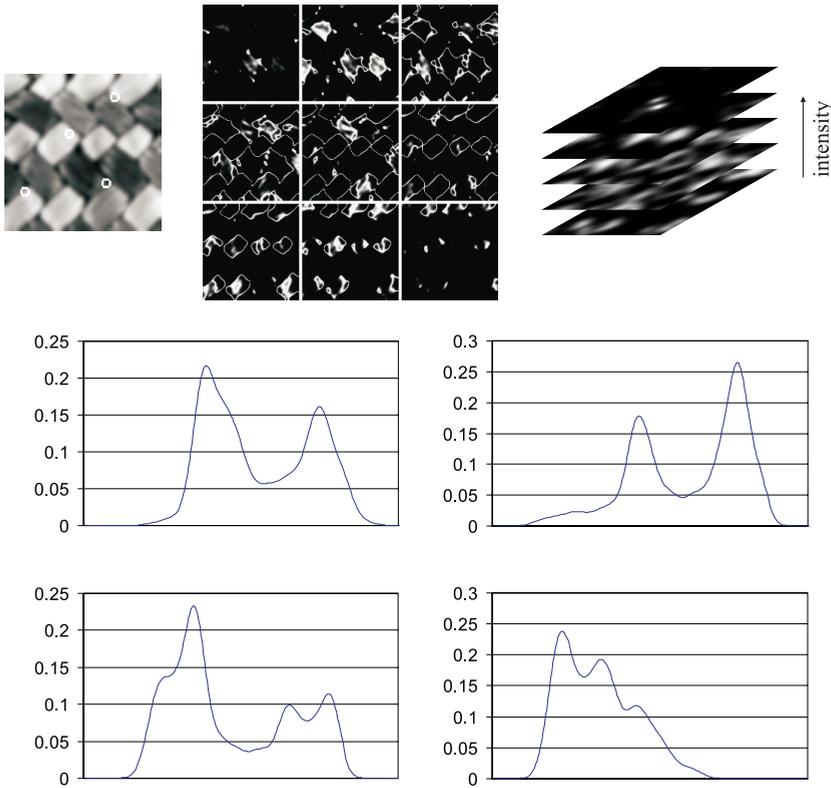


Figure 5.1: Illustration of the construction of a locally orderless image. The input is an image of 128×128 pixels, with intensities uniformly scaled between -1 and 1. Top row, left: this image is observed at scale $\sigma = 2$, with four locations indicated by small white circles. Top row, middle: for a given β , in this case 0.05, soft isophote images are computed. The panel displays 9 soft isophotes, with the intensity values ranging from -1 (top left) to 1 (bottom right). Top row, right: these soft isophote images are stacked together and observed at (blurred to) extent α , in this case 8, to form the locally orderless image. The extra dimension, in the z -direction is intensity. This schematic illustration shows only a few of the stacked slices. Bottom: for each location, the vertical direction in the locally orderless image represents a histogram with bin-width β for a Gaussian region of interest of width α . The panel shows four histograms, for the four locations indicated with circles.

where D is the dimensionality of the input image $L(\mathbf{x})$. Note that $H(\mathbf{x}, i; \sigma, \beta, \alpha)$ has a dimensionality 1 higher than that of $L(\mathbf{x})$. This is the result of the fact that scalar pixel values (dimensionality 0) have been replaced with marginal histograms (dimensionality 1). The construction is called “locally orderless” because at each location the pixel values have been replaced by orderless sets, *viz.* histograms. Figure 5.1 illustrates the construction of a locally orderless image.

Filter banks

So far, we have not specified the input image L , except that it can be observed at any scale σ . This implies that we take as input for the construction of the locally orderless image a family of images, obtained by filtering the original image with Gaussians of various scales σ . For texture classification it may be interesting to consider the response of the image to a larger class of filters than just Gaussians. In general, we can introduce a set of filters f , which can always be scaled by a parameter σ . In practice, the filters f will be a countable set. We indicate them with $f_1 \dots f_n$.

In this work we will not give a thorough discussion on how to construct a filter bank that is optimally tuned for the task of texture feature extraction, although this is an interesting and non-trivial matter. In the context of LOIs, which are a natural extension of the linear scale-space theory of image observations to histograms, the N-jet of derivatives up to a given order appears the natural choice for a filter bank [137,70], analogous to the use of a Taylor expansion to locally describe a function. This type of filter banks will be used, where the n -th order derivative in the direction θ will be indicated as L_n^θ . Notice that these derivatives are very similar to Gabor filters for higher orders. They can be “steered”, that is, computed for arbitrary θ for a given order n from the cross-derivatives of that order [73].

These filter banks are obviously sensitive to orientation and gray level transformations. The use of Gaussian receptive field families allows one to directly use results of scale-space theory to construct filters which are invariant to rotations [71] and gray-level transformations [69], if desired.

Joint probability distributions

The histograms obtained from Eq. 5.3 are marginal, *i.e.* a function of a single intensity i . General statistics can be obtained by considering a number of intensities from locations with a fixed spatial relation, that we can indicate as a vector $\mathbf{i} = i_1 \dots i_j$. In this way we obtain j -dimensional histograms or joint probability distributions. There may exist a meaningful order in the set of filtered images, *e.g.* when the set is a scale space or a steered orientation bundle. In that case it could make sense to consider intensities in different images. As in [267], the tuple of locations $i_1 \dots i_j$ are the vertices of an j -polygon in the set of filtered images (see Figure 5.2).

If the polygon consists of a single vertex ($j = 1$), we are dealing with a filter bank method. The difference with a classical filter bank method is that we have explicitly defined the construction as a function of the scale parameters σ , β and α . We can use this to perform different types of multi-resolution texture analysis. In the next section, we will evaluate the performance of such a multi-resolution filter bank system.

Haralick’s classical co-occurrence matrices framework [101] is obtained by using a polygon that consists of a single line segment ($j = 2$), with only the raw image taken as input, and α put to ∞ , and β fixed. If the scale space $L(\mathbf{x}; \sigma)$ is taken as input, the result is a multi-resolution version of the co-occurrence matrices parameterized by σ , β and α . The performance of such a system for multiple values of σ , β and the line segment (distance and relative orientation of the two locations) will be evaluated in the next section.

The combination of a filter bank and two locations within a single filtered image has been investigated as generalized co-occurrence matrices by Davis [54]. Oja and Valkealahti [192] [244] have presented results using an j -polygon on only the raw image as input.

Feature extraction

Although the raw histograms can be used for analysis (texture classification, as in [247] or texture generation as in [280]), usually features are extracted from the histograms. The most obvious and most widely followed approach is to use moments [187]. For marginal histograms $H(i)$, the first 4 moments around the mean, m_1, m_2, m_3, m_4 , are usually defined as

$$m_1 = \int iH(i)di \quad (5.4)$$

$$m_2 = \sqrt{\int (i - m_1)^2 H(i)di} \quad (5.5)$$

$$m_3 = \frac{1}{m_2^3} \int (i - m_1)^3 H(i)di \quad (5.6)$$

$$m_4 = \frac{1}{m_2^4} \int (i - m_1)^4 H(i)di - 3 \quad (5.7)$$

These moments are denoted as mean, standard deviation, skew and kurtosis, respectively. The skew is an indication of the symmetry of the distribution. In this definition, the kurtosis is 0 for a Gaussian distribution and indicates how much the distribution deviates from a Gaussian.

For joint probability distributions, moments could be used as well. However, for feature extraction from co-occurrence matrices commonly some of the 14 measures proposed by Haralick [101] are used. Most often 5 features are considered, referred to as energy (EG), entropy (EP), local homogeneity (LH), correlation (CO) and inertia (IN). If $H(i_1, i_2)$ denotes the two-dimensional histogram of two intensities i_1 and i_2 they are defined as

$$\text{EG} = \iint H(i_1, i_2)^2 di_1 di_2 \quad (5.8)$$

$$\text{EP} = - \iint H(i_1, i_2) \log H(i_1, i_2) di_1 di_2 \quad (5.9)$$

$$\text{LH} = \iint \frac{H(i_1, i_2)}{1 + (i_1 - i_2)^2} di_1 di_2 \quad (5.10)$$

$$\text{CO} = \iint \frac{(i_1 - \mu_{i_1})(i_2 - \mu_{i_2})H(i_1, i_2)}{\sigma_{i_1}\sigma_{i_2}} di_1 di_2 \quad (5.11)$$

$$\text{IN} = \iint (i_1 - i_2)^2 H(i_1, i_2)^2 di_1 di_2, \quad (5.12)$$

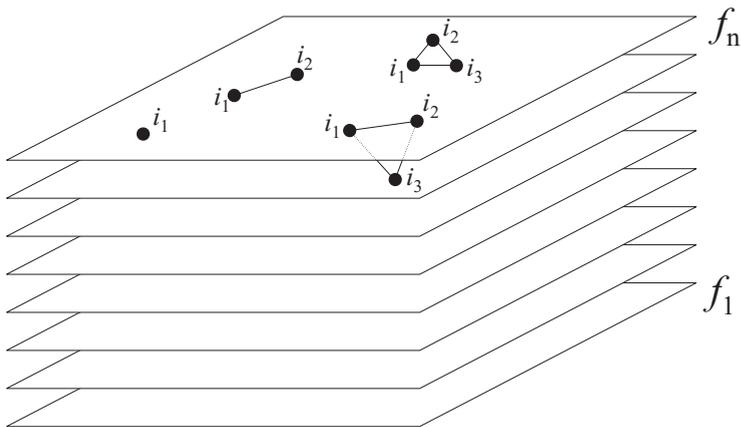


Figure 5.2: Multi-dimensional histograms are obtained by considering locations $i_1 \dots i_j$ with fixed spatial relations defined by the vertices of a polygon that is moved along each location in the set of filtered images $f_1 \dots f_n$. In the illustration the case of a single location is shown (only i_1), and examples of two (a line segment from i_1 to i_2) and three locations. Usually, the locations will lie within a single filtered image and the analysis will be performed for each filtered image separately (the first example of the triangle (i_1, i_2, i_3)). But if a meaningful order of the filtered images exist, it may also be useful to consider locations in different images (the second example of the triangle (i_1, i_2, i_3)).

where μ and σ denote the first and second moment around the mean of the joint distribution,

$$\mu_{i_1} = \iint i_1 H(i_1, i_2) di_1 di_2 \quad (5.13)$$

$$\mu_{i_2} = \iint i_2 H(i_1, i_2) di_1 di_2 \quad (5.14)$$

$$\sigma_{i_1} = \iint (i_1 - \mu_{i_1})^2 H(i_1, i_2) di_1 di_2 \quad (5.15)$$

$$\sigma_{i_2} = \iint (i_2 - \mu_{i_2})^2 H(i_1, i_2) di_1 di_2. \quad (5.16)$$

In the experiments in the next section, we will use moments for marginal distributions and these five features for co-occurrence matrices.

Dimensionality and relations between scale parameters

The proposed generalized locally orderless images are very high dimensional objects. Their dimensionality is given by the sum of the dimension of the input image and that of the intensity vector \mathbf{i} , plus the three scale parameters, possibly parameters that

vary the j -polygon of spatial relations for the locations that make up \mathbf{i} , and, in case of the filter bank of steered derivatives, the order of derivatives n and the orientation θ . Obviously, sparse sampling of some of these dimensions is required for practical applications.

Rules of thumb may be derived by considering the specific roles played by various parameters in the framework.

The order of derivative n determines the type of structure that is extracted by the filter. $n = 0$ detects blobs, $n = 1$ responds on edges, $n = 2$ is sensitive to line-structures. The orientation θ obviously tunes the orientation of the detected texture primitives. The number of independent filters per order is equal to $n + 1$.

The inner scale σ of the applied filters tunes the size of the primitives that are detected. Using various values of σ allows one to distinguish textures of primitives with similar shape but different sizes.

To characterize texture a description at multiple scales is essential. Haralick [100] put it as follows: “to characterize texture, we must characterize the tonal primitive properties as well as the spatial interrelationships between them. This implies that texture is really a two-layered structure, the first layer having to do with specifying the local properties which manifest themselves in tonal primitives and the second layer having to do with specifying the organization among the tonal primitives.” This clarifies the role of the extent α . While the filters parameterized by n, θ, σ extract specific tonal primitives (or “textons”, a term due to Julesz [123,122]), the distribution within an extent α describes the placement rules or spatial interrelationships between the primitives. Thus a multi-resolution analysis for α can distinguish textures with similar primitives but different placement rules.

The tonal scale β has the dimension of intensity. Thus it plays a role similar to σ but in the intensity domain. Varying β is a way to analyze the structure within the histograms or co-occurrence matrices. If two texture patches are identical except for a non-linear gray level transformation that has reduced the contrast between the primitives in a certain intensity range, a multi-resolution analysis for β may distinguish between such texture pairs.

Although the three scale parameters σ, α, β are of a fundamentally different character, there are related in practice. It does not make sense to consider situations where $\alpha \ll \sigma$ because below the inner scale of an observation no image structure is discernible and a histogram of a region of interest that is smaller than this scale is obviously a peak. Increasing α always increases the complexity of the histograms. In the case where $\alpha \rightarrow \infty$ the histogram no longer depends on \mathbf{x} which is of course a very efficient way to reduce the dimensionality of the locally orderless images. The dimensionless number $(\alpha/\sigma)^2$ is the logon content, a measure for the amount of information that is present at each location of the LOI. One may fix this value by coupling σ and α as Gårding and Lindeberg did [76]. Increasing β always reduces the complexity of the histogram.

Whether the use of multiple scales improves texture classification depends also on the features extracted from LOIs. In particular, when moments are used as features, the tonal scale β is immaterial. This can be seen by considering the expression for m_n , the n th moment around the mean m :

$$m_n = \int (i - m)^n H(i) di. \quad (5.17)$$

This expression is linear in $H(i)$, which means that if we want to investigate the effect of blurring $H(i)$ on the moment, we may do this by considering what happens at a single value i' .

If we use a blurred histogram $H(i; \beta)$ we have instead of a single value at i' a Gaussian with width β :

$$m_n = \int (i - m)^n \frac{1}{\beta\sqrt{2\pi}} \exp\left(\frac{-(i - i')^2}{2\beta^2}\right) di, \quad (5.18)$$

For $\beta \rightarrow 0$ this reduces to $(i - m)^n$. For $\beta > 0$, the 0th and 1st order moment remain equal. Higher order moments change because equal amounts of weight are moved towards the mean and away from it and the change in contribution of these two parts to the moment are different. The expression for the general integral in (5.18) is cumbersome to express, so we state the results for the first few moments:

$$\begin{aligned} m_0 &= 1, \\ m_1 &= (i' - m), \\ m_2 &= (i' - m)^2 + \beta^2, \\ m_3 &= (i' - m)^3 + 3\beta^2(i' - m), \\ m_4 &= (i' - m)^4 + 6\beta^2(i' - m)^2 + 3\beta^4, \end{aligned}$$

and so on. Thus by increasing β , m_0 and m_1 are unchanged and the higher order moments are simply increased by fixed amounts proportional to $H(i')$ and earlier moments. Therefore, if we use moments as features, using multiple values of β does not yield extra information.

Experiments

We tested several texture feature sets on a single large data set. The aim of these experiments is to evaluate the performance of the texture feature schemes, to compare between filter bank methods and co-occurrence features and especially to investigate whether multi-resolution approaches for the different scale parameters in the extended LOIs are effective in practice.

Data

We collected texture patches from a large variety of natural texture images. These images originated from the VisTex database, maintained by the Vision and Modeling

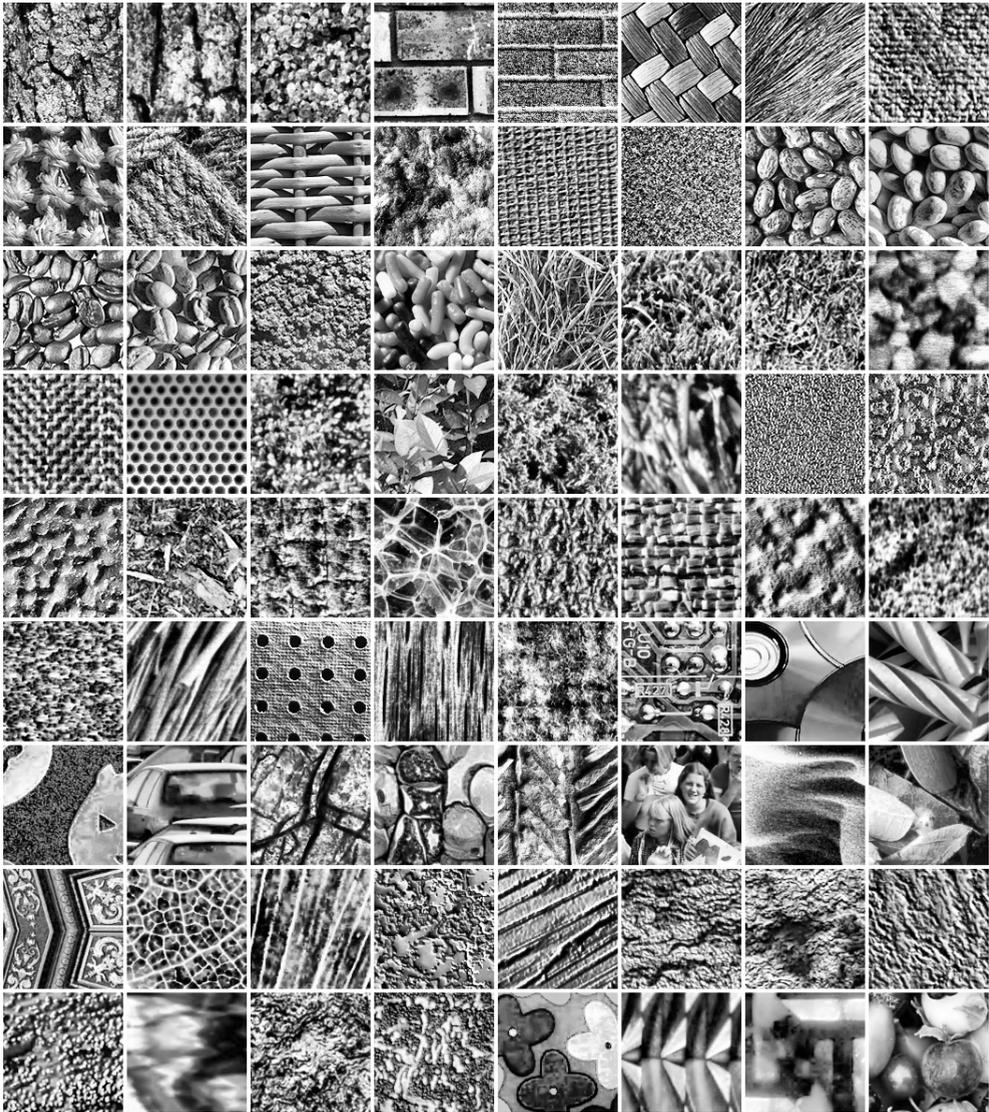


Figure 5.3: The 72 different textures used in the experiments. One patch for each class is shown; the total set contains 16 patches of each class, 1152 textures in total. See figure 5.4 for a display with all samples.

group at the MIT Media Lab [248], the image collection distributed with the Corel-Draw software suite from Corel, and the SIPI database [255]. Most images in the SIPI database are taken from the Brodatz collection [26].

We selected all those images from these databases that contained large textures from which 16 (4 by 4) smaller texture patches could be cropped in such a way that

each cropped patch was still representative of the texture shown in the image as a whole. Consequently, texture images with global gradients, *e.g.* perspective views of a water surface, were excluded.

The data set includes a wide variety of textures from natural scenes and man-made materials. One instance of each class is shown in Figure 5.3 and all patches are shown in Figure 5.4. Note that the selection contains several classes that are very much alike. For example, the first two classes are both tree bark, three classes of grass are included and two sets of two classes are cropped from identical scenes, photographed with different illumination conditions.

The total number of images is 72, each with a resolution of 512×512 pixels and 256 gray levels. Each image was subdivided into 16 (4 by 4) adjacent but non-overlapping patches of 128×128 pixels. The gray level probability (histogram) of each patch was equalized to eliminate difference in first order statistics between the patches, and scaled to the range $[-1, 1]$. The total number of sample patches measures $72 \times 16 = 1152$. To the best of our knowledge, this data set contains more different classes and samples than sets used in previously published studies.

Classifier

We used a k -nearest-neighbor (k NN) classifier and the leave-one-out method for all experiments. Before classification, each feature was normalized to zero mean and unit variance over all samples in the set. The k NN classifier has several advantages over alternatives such as neural networks. No training is required and the results are repeatable. There is only one tuneable parameter, k . We used a fast implementation of the k NN classifier by Arya and Mount [8], available on the web at <http://www.cs.umd.edu/~mount/ANN>. We use sequential feature forward selection (Whitney's method) [264, 203, 223]. This means we start with 0 features. All available features are added one by one and the performance is computed, whereupon the feature that yields the highest gain in performance (or the smallest decrease, in case adding any feature decreases performance) is added to the set. This procedure is repeated until a given number of features, or all features, have been added. We selected at most 40 features. Note that a set of n features obtained in this way does not necessarily correspond to the best subset of n features [75]. In principle, one must be careful when a data set is used for both feature selection and evaluation when the number of features is large compared to the number of classes and the number of samples. It is possible in those cases that some (random) features perform well by chance. In this case, however, with 1152 samples from 72 classes, this is unlikely, even for the large number of features considered for some systems in this study. From an application's point of view it could be useful to consider (linear) combinations of features. We choose not to do this, in order to be able to identify "good" features directly.

We performed the classification with $k = 1, 3, 5$. In general performance was best with $k = 1$ and slightly decreased for $k = 3$ and $k = 5$. Therefore we show results of $k = 1$.

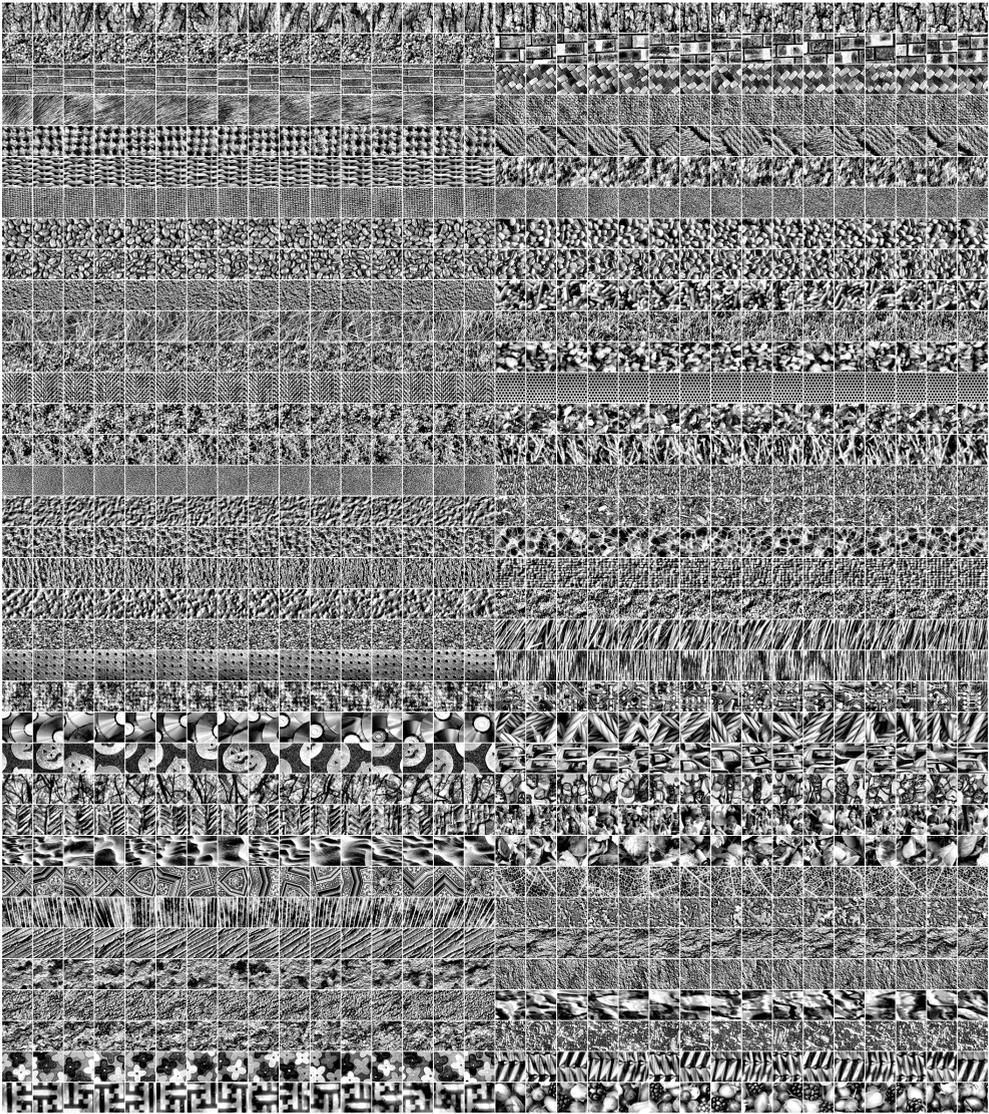


Figure 5.4: The complete set of samples used in the experiments. The total number is of samples is 1152, from 72 classes, 16 images per class. Two classes are shown per row. A larger example of one sample from each class is shown in [Figure 5.3](#).

Texture feature sets

In order to compare the effect of varying the scale parameters in our framework, we added two classical methods.

The first is the 3×3 Laws filter bank [150], which is representative of filter bank methods based on well known transforms such as the Hadamard transform, the discrete sine and cosine transform etc. [241]. The 3×3 Laws filters are constructed from 3-tap blob, edge and spot filters ($L3 = |1\ 2\ 1|$, $E3 = |1\ 0\ -1|$, $S3 = |1\ -2\ 1|$). For instance, the notation L3S3 stands for 3×3 kernel obtained by applying the L3 filter in the x -direction and the S3 filter in the y -direction. The second classical method uses features extracted from co-occurrence matrices, using the five features defined by Eqs. 5.8–5.12. We extend the Laws filter bank with multiple resolutions for α . We compare this with a filter bank based on Gaussian derivatives. This filter bank is extended to a multi-resolution approach for both σ and α . The co-occurrence scheme is extended to multiple resolutions for σ , β and multiple separation distances d between the two points.

A detailed description of each system is given below.

1. Classical 3×3 Laws filter bank. As such, $\sigma = 0$ and $\alpha = \infty$. We use the second, third and fourth moment of the histogram. Using the 1st moment makes no sense because the input images have equal histograms. There are 9 filters. Thus we have 27 features in total.
2. Classical 3×3 Laws filter bank using multi-resolution for extent α . The system is similar to the classical 3×3 Laws filter, except that the cases $\alpha = 1, 2, 4$ are added. Thus we have a total of 104 features.
3. Filter bank of Gaussian derivatives. Similar to the classical 3×3 Laws filter bank, except that a filter bank of Gaussian derivatives is used at the fixed scale $\sigma = 1$. Filters used are $L_0, L_1^{0^\circ}, L_1^{90^\circ}, L_2^{0^\circ}, L_2^{60^\circ}, L_2^{120^\circ}$. The total number of features is 18.
4. Filter bank of Gaussian derivatives using multiple resolutions for the inner scale σ . Similar to the previous filter bank of Gaussian derivatives except that this system uses three scales $\sigma = 0.5, 1, 2$. Thus we have 54 features.
5. Filter bank of Gaussian derivatives using multiple resolutions for the inner scale σ and extent α . The previous system is now extended to a multi-resolution approach for α . As in system 2, we consider $\alpha = 1, 2, 4, \infty$. The total number of features is therefore 216.
6. Classical co-occurrence features. As such $\sigma = 0$ and $\alpha = \infty$. For the size of the co-occurrence matrices we choose 256. This corresponds to $\beta = 0$ because the images all had 8 bit intensity values originally. We use the 5 features defined in Eqs. 5.8–5.12 (energy, entropy, local homogeneity, correlation and inertia) for 4 sets of pixels with separation distance $d = 1$ at $0^\circ, 45^\circ, 90^\circ, 135^\circ$. Thus we have 20 features.
7. Classical co-occurrence features with multiple resolutions for the inner scale σ . Similar to the previous system, but now we use $\sigma = 0, 0.5, 1, 2$. Thus we have 80 features in total.

8. Classical co-occurrence features with multiple resolutions for the tonal scale β . Similar to the classical co-occurrence feature system, but now we use $\beta = 0, 1, 2, 4, 8$. Thus we have 100 features in total.
9. Classical co-occurrence features with multiple separation distances d . Similar to the classical co-occurrence feature system, but now we use $d = 1, 2, 4$. Thus we have 60 features in total.
10. Classical co-occurrence features with multiple separation distances d and multiple resolutions for the inner scale σ and tonal scale β . This combines the 3 previous systems. We vary both $\sigma = 0, 0.5, 1, 2$, $d = 1, 2, 4$, and $\beta = 0, 1, 2, 4, 8$. Thus we have 1200 features.

	result best set	# features	result whole set	# features
System 1	0.9748	19	0.9640	27
System 2	0.9835	37	0.8446	108
System 3	0.9409	15	0.9375	18
System 4	0.9869	40	0.9713	54
System 5	0.9922	38	0.9817	216

Table 5.1: Results for all filterbank systems (System 1 to 5) for the best combination of features (up to 40 features) and the complete set with all features. The number of features for the best subset up to 40 and for the whole set is also given.

Results and discussion

For each system, we give the classification performance in terms of the part of correctly classified texture patches after 1 to 15 features have been added in Tables 5.2 and 5.5. We also give the performance of the best feature set after at most 40 features have been added, and the performance of the system with all features (Table 5.1 and 5.4). Table 5.3 and 5.6 explicitly list the best 10 features of each system. In Figure 5.5 the performance of each system is plotted as a function of the number of features.

The key result of the experiments is that all types of multi-resolution analysis increase the classification performance. We consider three different types of multi-resolution features extraction by varying σ , α and β respectively, while previous work on multi-resolution texture analysis was limited to σ only.

The effect of using multiple extents α can be seen by comparing system 1, classical Laws filters, with system 3, the same filters for multiple values of α . The effect is even more clear by comparing system 4, a filter bank system with multiple resolutions σ , with system 5, the same system extended with multiple resolutions for α . In both cases there is a significant increase in performance.

Likewise, the effect of multiple inner scales σ can be seen from comparing system 3 with system 4 and also system 6 with system 7. Using multiple resolutions for the inner scale σ has a large performance benefit.

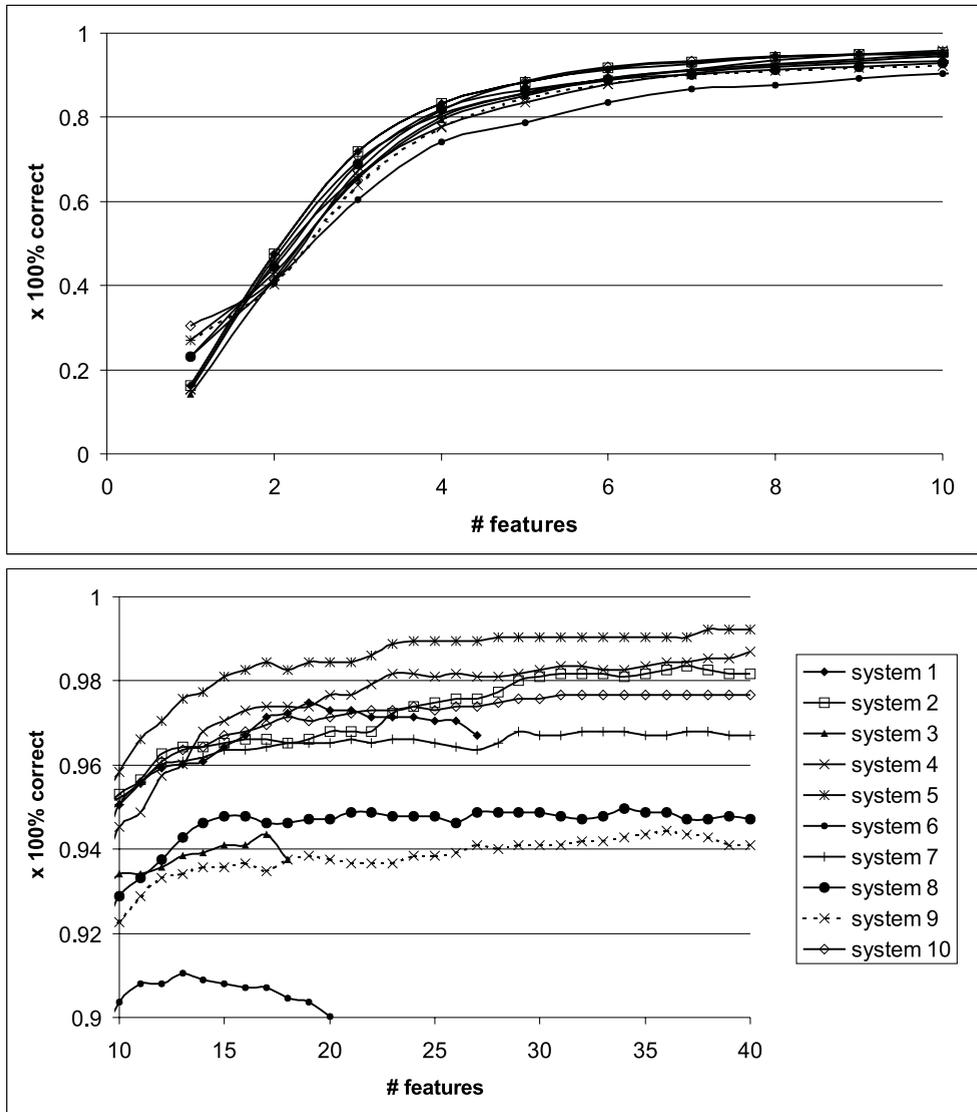


Figure 5.5: Classification result of each system as a function of the number of selected features. For display purposes, the graph is split in two parts. The top graph shows the result for 0 to 10 features, the bottom graph displays the results for 10 to 40 features.

# features	System				
	1	2	3	4	5
1	0.1623	0.1623	0.1423	0.1527	0.1527
2	0.4756	0.4756	0.4157	0.4539	0.4652
3	0.7178	0.7178	0.6553	0.6614	0.6953
4	0.8324	0.8324	0.7934	0.7777	0.8081
5	0.8828	0.8828	0.8506	0.8350	0.8567
6	0.9123	0.9175	0.8923	0.8793	0.8888
7	0.9262	0.9322	0.9105	0.9045	0.9131
8	0.9444	0.9418	0.9192	0.9253	0.9366
9	0.9496	0.9487	0.9288	0.9331	0.9505
10	0.9505	0.9531	0.9340	0.9453	0.9583
11	0.9557	0.9565	0.9340	0.9487	0.9661
12	0.9592	0.9626	0.9357	0.9574	0.9704
13	0.9600	0.9644	0.9383	0.9600	0.9756
14	0.9609	0.9644	0.9392	0.9678	0.9774
15	0.9644	0.9652	0.9409	0.9704	0.9809

Table 5.2: Cumulative classification results for filter bank systems (System 1 to 5). The best 10 features are listed in Table 5.3.

#	System 1		System 2			System 3		System 4			System 5			
	filter	m	filter	m	α	filter	m	filter	m	σ	filter	m	σ	α
1	E3S3	2	E3S3	2	∞	L	2	L	2	0.5	L	2	0.5	∞
2	L3E3	2	L3E3	2	∞	$L_1^{0^\circ}$	2	L	2	2	L	2	0.5	1
3	S3S3	4	S3S3	4	∞	$L_1^{90^\circ}$	2	$L_1^{90^\circ}$	4	0.5	$L_2^{0^\circ}$	2	1	∞
4	S3L3	2	S3L3	2	∞	$L_2^{120^\circ}$	4	L_2^0	2	1	$L_1^{90^\circ}$	4	0.5	4
5	L3S3	3	L3S3	3	∞	$L_2^{60^\circ}$	2	$L_2^{60^\circ}$	4	0.5	$L_1^{90^\circ}$	4	1	2
6	E3E3	2	L3L3	2	1	$L_2^{60^\circ}$	3	$L_1^{60^\circ}$	4	2	$L_2^{60^\circ}$	3	1	1
7	E3L3	2	L3E3	4	∞	$L_2^{60^\circ}$	4	$L_2^{60^\circ}$	2	1	$L_1^{60^\circ}$	3	0.5	1
8	L3E3	4	E3E3	2	∞	$L_1^{0^\circ}$	4	L	4	1	$L_2^{60^\circ}$	4	1	∞
9	L3S3	2	L3E3	2	1	$L_2^{120^\circ}$	2	L_2^0	2	0.5	$L_2^{120^\circ}$	2	0.5	∞
10	L3S3	2	L3L3	4	∞	L_2^0	4	L	3	0.5	L_2^0	3	0.5	1

Table 5.3: Best 10 features for filter bank systems (System 1 to 5). The corresponding classification results are listed in Table 5.2.

	result best set	# features	result whole set	# features
System 6	0.9105	13	0.9001	20
System 7	0.9678	20	0.8368	100
System 8	0.9496	34	0.9505	80
System 9	0.9444	36	0.9227	60
System 10	0.9765	31	0.9045	1200

Table 5.4: Classification results for co-occurrence systems (System 6 to 10) for the best combination of features (up to 40 features) and the complete set with all features. The number of features in each set and subset is also given.

Finally, comparing systems 6 and 8 shows that considering multiple tonal scales β boosts the performance of texture classification with co-occurrence matrices. This is an interesting result since β has always been kept at a single fixed value in previous studies.

The three multi-resolution approaches are truly different: combined they lead to the most effective systems for texture classification. System 5 obtains the overall highest score with 0.9922 for 38 features. In this case 9 out of 1152 texture patches were classified incorrectly. The fact that the various filters, features, moments and scales do indeed capture different aspects of texture is also supported by Tables 5.3 and 5.6; in almost all systems the best ten features include all types of filters, all moments, and each of the scales considered.

The experiments here indicate that filter bank systems outperform co-occurrence systems. This may be partly due to the fact that the 5 features computed from the co-occurrence matrices do not capture all the information they contain.

The errors made during classification may be partly attributed to “overlap” between classes in the sense that it is for example difficult to distinguish samples from different classes of grass from each other. Another cause of error is that some classes contain samples that are not representative for their class because they contain only a few textons. Examples can be found in the patches with faces, CDs and candlesticks. Human observers would probably classify these samples by recognizing the objects, using knowledge about the structure of these objects. All texture methods considered here are based on local statistics and are not capable of such high-level reasoning.

From Tables 5.1 and 5.4 results with and without feature selection can be compared. Feature selection boosts performance in nearly all cases. But in both cases, system 5 has the best performance and the ranking of systems hardly changes when feature selection is applied. An exception is system 10, which contains many features (more than the number of images) and therefore shows a large improvement in performance when the best features are selected. Another exception is the multi-resolution Laws set (system 2) which apparently contains many weak features. This indicates that some Laws filters are not very effective in distinguishing the classes in our dataset.

The particular choices we made for the filter bank, co-occurrence features, moments, number of scales to consider, and classification method, are inevitably some-

what arbitrary. It may well be possible that some other combination improves the results in particular applications.

# features	System				
	6	7	8	9	10
1	0.2300	0.2690	0.2300	0.2690	0.3038
2	0.4088	0.4305	0.4444	0.4027	0.4149
3	0.6041	0.6588	0.6892	0.6388	0.6736
4	0.7421	0.8012	0.8194	0.7760	0.8168
5	0.7881	0.8576	0.8654	0.8446	0.8862
6	0.8350	0.8914	0.8897	0.8793	0.9201
7	0.8663	0.9131	0.9019	0.9019	0.9340
8	0.8767	0.9262	0.9140	0.9105	0.9453
9	0.8932	0.9383	0.9192	0.9175	0.9505
10	0.9036	0.9505	0.9288	0.9227	0.9522
11	0.9079	0.9557	0.9331	0.9288	0.9557
12	0.9079	0.9600	0.9375	0.9331	0.9609
13	0.9105	0.9609	0.9427	0.9340	0.9635
14	0.9088	0.9618	0.9461	0.9357	0.9644
15	0.9079	0.9635	0.9479	0.9357	0.9670

Table 5.5: Cumulative classification results for co-occurrence systems (System 6 to 10) for the first 15 features. The best 10 features are listed in Table 5.6.

In the filter bank method β was not used as multi-resolution parameter because this does not provide extra information. For the co-occurrence matrices it is possible to also include α in the multi-resolution approach. However, this would be much more computationally expensive. When α is not fixed at ∞ , three extra dimensions appear in the framework (x, y and α itself). Therefore we have refrained from this multi-resolution analysis.

The framework could be used for texture segmentation as well. In this case, α may be used in the same way as it has been here – to serve as the extent over which to collect information regarding the placement rules of the textons. Alternatively, α may be used as region of interest over which to collect texture information, what has been called *integration scale* [260, 76].

Conclusions

We can summarize the results by stating the advantages of using generalized locally orderless images to derive texture features. It provides a framework in which all relevant scale parameters are explicitly identified instead of introduced in an ad hoc fashion. One can directly apply existing scale-space methods and perform a multi-resolution analysis that takes all scale parameters into account. This allows for the systematic extraction of texture features and generalizes existing methods. Our experiments suggest that all types of multi-resolution analysis are beneficial to classification

#	System 6	System 7		System 8		System 9	System 10		
	d	d	σ	d	β	d	d	σ	β
1	EP (1,1)	EP (1,1)	0.5	EP (1,1)	0	EP (-4,4)	EP (0,4)	0.5	0
2	EP (0,1)	EG (0,1)	0	LH (1,0)	4	IN (1,0)	EG (0,1)	0	4
3	EP (1,0)	EG (1,1)	2	LH (-1,1)	4	CO (-2,2)	EG (1,0)	0	2
4	LH (-1,1)	LH (1,0)	0.5	EG (0,1)	2	LH (-1,1)	LH (-1,1)	0	0
5	LH (1,0)	CO (1,0)	0	EG (1,0)	2	IN (4,0)	EG (2,0)	1	2
6	IN (0,1)	CO (1,0)	1	EG (1,1)	1	CO (0,2)	LH (1,1)	0	4
7	LH (1,1)	LH (1,1)	0	LH (0,1)	0	LH (1,1)	EG (0,1)	2	1
8	CO (1,1)	EG (0,1)	2	CO (1,0)	0	LH (1,0)	LH (-1,1)	0.5	8
9	CO (1,0)	LH (-1,1)	0	LH (1,1)	8	EG (0,1)	LH (1,0)	0	8
10	IN (-1,1)	EG (1,0)	1	LH (1,0)	0	IN (0,4)	LH (1,1)	0	0

Table 5.6: Best 10 features for co-occurrence systems (System 6 to 10). The Haralick features are abbreviated: EP = entropy, EG = energy, LH = local homogeneity, IN = inertia, CO = correlation. The corresponding classification results are listed in Table 5.5.

performance.

A system of 6 basis filters (zeroth order derivative, first order derivatives at two orientations and second order derivatives at three orientations) at three inner scales and four extents from which the second to fourth moment were used as features achieves, after feature selection, excellent performance (99.21%) on a data set of 1152 texture samples from 72 different classes.

Chapter 6

Active Shape Model Segmentation with Optimal Features

An active shape model segmentation scheme is presented that is steered by optimal local features, contrary to normalized first order derivative profiles, as in the original formulation by Cootes and Taylor [49, 48]. A non-linear k NN-classifier is used, instead of the linear Mahalanobis distance, to find optimal displacements for landmarks. The optimal features are automatically selected from the training images. The new approach produces significantly better results ($p < 0.001$) than the original scheme in four medical segmentation tasks: segmenting the right and left lung fields in chest radiographs and segmenting the cerebellum and corpus callosum in slices from MRI brain images.

THE ULTIMATE GOAL of computer vision is – as the word says – to make machines that can see. Given an image, in the form of an array of pixel values, the machine should be able to *segment* the image, *i.e.* divide it into meaningful parts, and locate known objects.

During the seventies and eighties, most researchers approached the segmentation problem in a *bottom-up* fashion: emphasis was on the analysis and design of filters that detect local structures such as edges, ridges, corners and T-junctions. The structure of an image can be described as a collection of such syntactical elements and their (spatial) relations, and such descriptions can produce segmentations. Unfortunately, these segmentations are often not very useful in practice. On the other hand, *top-down* or *model-based* or *active* approaches to segmentation were used successfully in highly constrained environments, *e.g.* in industrial inspection tasks. Often these methods are some form of template matching. Templates incorporate knowledge about both the shape of the object to be segmented and its appearance in the image, and are matched by correlation or generalized Hough transform techniques. But template matching will likely fail for segmentation tasks with a lot of variation in the appearance of objects and background, such as real-life images and medical data, even if several scaled and rotated versions of the template are used.

Active contours or snakes, first proposed by Kass *et al.* [125] and reviewed in [177], have been heralded as a new paradigm for segmentation. It was the ability of snakes to deform freely instead of rigidly that spurred this enthusiasm. Nevertheless, snakes have two inherent limitations, which make them unsuitable for many medical segmentation tasks. No a priori knowledge about the shape to be segmented can be expressed except for adjusting parameters that regulate the smoothness of the contour, and the image structure at object boundaries is prescribed by letting the snakes attract to edges or ridges in the image. In practice object boundaries often do not coincide with an edge or a ridge.

To overcome these limitations, researchers experimented with hand crafted parametric models. An illustrative example is the work of Yuille *et al.* [279] in which a deformable model of an eye is constructed from circles and parabolic patches and a heuristic cost function is proposed for the appearance of the image inside and on the border of these patches.

There are two main problems with parametric models. First of all they are *dedicated*, that is, limited to a single application. In that sense they can be compared with the lung segmentation methods proposed in Chapter 3. Second, there is no proof that they are optimal. Therefore there exists a need for generic segmentation schemes that can be trained with examples as to acquire a model of the shape of the object to be segmented (with its variations) and the appearance of the object in the image (with its variations). These methods are prototype-based which makes it easy to adapt them to new applications by replacing the prototypes; they use statistical techniques to extract the major variations from the prototypes in an optimal way (subject to certain constraints and assumptions, obviously).

Several of such schemes have been proposed [119, 20]. In this work we focus on Active Shape Models (ASMs) put forward by Cootes and Taylor [49]. Note that the term “active shape models” is used in the literature in a more general sense, as a shape model that includes shape variations. In our case it refers to the complete segmentation scheme (and not just the shape model) proposed in [49] and [48]. The latter reference is a technical report that describes the ASM segmentation method in detail; we have implemented the method based on this description. The shape model in ASMs is given by the principal components of vectors in which landmark points are stacked. The appearance model is limited to the border of the object and consists of the normalized first derivative of profiles sampled perpendicular to each landmark. The cost (or energy) function to be minimized is the Mahalanobis distance of these first derivative profiles. The fitting procedure is an alternation of landmark displacements and model fitting in a multi-resolution framework. The ASM scheme is described step-by-step in the next section.

Several comparable approaches are found in the literature. Shapes and objects have been modeled by landmarks, finite element methods, Fourier descriptors and by expansion in spherical harmonics (especially for surfaces in 3D [132, 95]). For shape models in general see [88, 63, 22]. Jain *et al.* [120] have presented a Bayesian framework in which templates are deformed and more probable deformations are more likely to occur. They use a coarse-to-fine search algorithm. Ronfard [207] has used statistics of object and background appearance in the energy function of a snake. Brejl and Sonka [25] have described a scheme similar to ASMs but with a non-linear shape and

appearance model that is optimized with an energy function after an exhaustive search to find a suitable initialization. Pizer *et al.* [200] describe an object model that consists of linked primitives which can be fitted to images using methods similar to ASMs. Cootes and Taylor have explored active appearance models (AAMs) as an alternative to ASMs. In AAMs a combined principal component analysis of the landmarks and pixel values inside the object is made which allows one to generate images. The iterative steps in the optimization of the segmentation are steered by the difference between the true pixel values and the modeled pixel values within the object. Sclaroff and co-workers [216, 215] have proposed a comparable method in which the object is modeled as a finite element model. While there are differences, the general lay-out of these schemes is similar: there is a shape model that ensures that the segmentation can only produce plausible shapes, there is an appearance model that ensures that the segmentation places the object at a location where the image structure around the border or within the object is similar to what is expected from the training images, and there is an algorithm for fitting the model by minimizing some cost function. Usually the algorithm is implemented in a multi-resolution fashion.

ASMs have been used for several segmentation tasks in medical images [47, 17]. Our contribution in this chapter is a new type of border appearance model. Instead of using the normalized first derivative profile, we consider a general set of local image structure descriptors based upon the framework of locally orderless images, introduced in Chapters 4 and 5. Subsequently a statistical analysis (feature selection with a k NN-classifier) is performed to learn which descriptors are the most informative at each resolution, and at each landmark.

We present experiments on synthetic images, chest radiographs, and brain MR data with both the original ASM scheme and the new method with optimal features. In the final sections the results are presented and discussed.

Active Shape Models

This section briefly reviews the ASM segmentation scheme. We follow the description and notation of [48]. The parameters in the scheme are listed in Table 6.1.

Shape Model

An object is described by n points, referred to as landmark points. The landmark points are (manually) determined in a set of s training images. From these collections of landmark points, a point distribution model [63] is constructed as follows. The landmark points $(x_1, y_1), \dots, (x_n, y_n)$ are stacked in shape vectors

$$\mathbf{x} = (x_1, y_1, \dots, x_n, y_n)^T. \quad (6.1)$$

These shapes can be aligned by translating, rotating and scaling them so as to minimize the sum of squared distances between the landmark points. An iterative scheme known as Procrustes analysis [86] is used to align the shapes which includes a projection into tangent space of each shape, see [48] for details. However, this

General	
s	number of training images from which shape and appearance model are computed
Shape model	
n	number of landmark points
alignment	true or false, whether or not to align the shapes before computing the shape model
t	number of modes in the shape model (controlled by f_v)
f_v	part of variance to be explained by the shape model, determines number of modes (0.98 with alignment, 0.995 without alignment)
m	bounds on eigenvalues λ_i (2.0)
Appearance model	
k	number of points in profile on either side of the landmark point, giving profiles of length $2k + 1$ (2)
N_{grid}	(<i>optimal features method only</i>) size of $N_{grid} \times N_{grid}$ grid of points from which features are sampled in training images (5)
f_{max}	(<i>optimal features method only</i>) maximum number of features to select (10)
k_{NN}	(<i>optimal features method only</i>) number of neighbors used in k NN classifier during searching and feature selection (5)
Search algorithm	
n_s	number of new positions to evaluate during each iteration on either side of current landmark position; total number of positions evaluated is $2n_s + 1$ (4)
L_{max}	number of resolution levels (4)
N_{max}	number of iterations per resolution level (10)

Table 6.1: Parameters for active shape models (original scheme and new method with optimal features). Between parentheses the values used in the experiments in this chapter are given.

alignment can also be omitted. This results in a shape model that can generate only shapes with a size, position and orientation that is consistent with the supplied examples. We have found that if enough training data is available, an unaligned model can lead to better performance of the segmentation algorithm.

Principal component analysis (PCA) is applied to the (aligned or not aligned) shape vectors \mathbf{x} by computing the mean shape,

$$\bar{\mathbf{x}} = \frac{1}{s} \sum_{i=1}^s \mathbf{x}_i, \quad (6.2)$$

the covariance

$$\mathbf{S} = \frac{1}{s-1} \sum_{i=1}^s (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \quad (6.3)$$

and the eigensystem of the covariance matrix. The eigenvectors corresponding to the t largest eigenvalues λ_i are retained in a matrix $\Phi = (\phi_1 | \phi_2 | \dots | \phi_t)$. A shape can now be approximated by

$$\mathbf{x} \approx \bar{\mathbf{x}} + \Phi \mathbf{b} \quad (6.4)$$

where \mathbf{b} is a vector of t elements containing the model parameters, computed by

$$\mathbf{b} = \Phi^T (\mathbf{x} - \bar{\mathbf{x}}). \quad (6.5)$$

When fitting the model to a set of points, the values of \mathbf{b} are constrained to lie in the range $\pm m\sqrt{\lambda_i}$ where m usually has a value between 2 and 3.

The number t of eigenvalues to retain is chosen so as to explain a certain proportion f_v of the variance in the training shapes, usually ranging from 90% to 99,5%. The desired number of modes is given by the smallest t for which

$$\sum_{i=1}^t \lambda_i \geq f_v \sum_{i=1}^{2n} \lambda_i. \quad (6.6)$$

Appearance Model

The appearance model that describes the typical image structure around each landmark is obtained from pixel profiles, sampled (using linear interpolation) around each landmark, perpendicular to the contour.

Note that this requires a notion of connectivity between the landmark points from which the perpendicular direction can be computed. The direction perpendicular to a landmark (x_n, y_n) is computed by rotating the vector that runs from (x_{n-1}, y_{n-1}) to (x_{n+1}, y_{n+1}) over 90° . In the applications presented here the objects are closed contours, so for the first landmark, the last landmark and the second landmark are used; for the last landmark, the one but last landmark and the first landmark are used.

On either side k pixels are sampled using a fixed step size, which gives profiles of length $2k + 1$. Cootes and Taylor propose to use the normalized first derivatives of these profiles to build the appearance model. The derivatives are computed using finite differences between the $(j - 1)$ th and the $(j + 1)$ th point. The normalization is such that the sum of absolute values of the elements in the derivative profile is 1.

Denoting these normalized derivative profiles as $\mathbf{g}_1 \dots \mathbf{g}_s$, the mean profile $\bar{\mathbf{g}}$ and the covariance matrix \mathbf{S}_g are computed for each landmark. This allows for the computation of the Mahalanobis distance between a new profile \mathbf{g}_i and the profile model

$$f(\mathbf{g}_i) = (\mathbf{g}_i - \bar{\mathbf{g}}) \mathbf{S}_g^{-1} (\mathbf{g}_i - \bar{\mathbf{g}}) \quad (6.7)$$

Minimizing this Mahalanobis distance $f(\mathbf{g}_i)$ is equivalent to maximizing the probability that \mathbf{g}_i originates from the distribution $\{\mathbf{g}_1 \dots \mathbf{g}_s\}$.

Multi-resolution framework

These profile models, given by $\bar{\mathbf{g}}$ and \mathbf{S}_g , are constructed for multiple resolutions. The number of resolutions is denoted by L_{max} . The finest resolution uses the original image and a step size of 1 pixel when sampling the profiles. The next resolution is the image observed at scale $\sigma = 1$ and a step size of 2 pixels. Subsequent levels are constructed by doubling the image scale and the step size. Note that we do not, as proposed by Cootes and Taylor, subsample the images.

Optimization algorithm

Shapes are fitted in an iterative manner, starting from the mean shape. Each landmark is moved along the direction perpendicular to the contour to n_s positions on either side, evaluating a total of $2n_s + 1$ positions. The step size is, again, $2^{(i-1)}$ pixels for the i th resolution level. The landmark is put at the position with the lowest Mahalanobis distance. After moving all landmarks, the shape model is fitted to the displaced points, yielding an updated segmentation. This is repeated a fixed number of N_{max} times at each resolution, from coarse to fine. We always make these N_{max} iterations, contrary to Cootes and Taylor who move to a finer resolution if a convergence criterion is reached before N_{max} iterations are made.

Improving ASMs

In this section some observations are made regarding ASMs and some proposed and some possible extensions and improvements of the method are discussed.

1. The shape model is fitted by projecting a shape in the Nn -dimensional space (N the dimension of the images, n the number of landmarks) upon the subspace spanned by the t largest eigenvectors and by truncating the model parameters \mathbf{b} so that the point is inside the box bounded by $\pm m\sqrt{\lambda_i}$. Thus there is no smooth penalty term for unlikely shapes; all shapes in the box are allowed, outside the box no shape is allowed. Clearly this can be refined in many ways.
2. The shape model uses PCA and therefore assumes that the distribution of shapes \mathbf{x} in the Nn -dimensional space is linear. If this is not true, non-linear models, such as mixture models, could be more suitable (see for example [224]).
3. By projecting a shape according to Eq. (6.5), *i.e.* fitting the shape model, the resulting model parameters \mathbf{b} minimize the sum of squared distances between true positions and modeled positions. In practice it can be desirable to minimize

only the distance between true and model positions in the direction perpendicular to the object contour because deviation along the contour does not change whether pixels are inside or outside the object. In [105] it is demonstrated how to perform this projection on the contour.

4. The original version of the appearance model is always based on normalized first derivative profiles. There is no a priori reason why this should be an optimal choice. In this chapter we propose an alternative to the border appearance model that uses profiles obtained from optimal features, derived automatically from training images.
5. The Mahalanobis distance in Eq. (6.7) assumes a linear distribution of profiles. In practice the distributions of profiles will often be non-linear, for example in cases where the background of the object may be one of several possible choices. The ASM scheme proposed in this chapter uses a non-linear classifier in the appearance model and can therefore deal with non-linear distributions.
6. Because of the multi-resolution implementation, the initial position of the object generally does not have to be very precise, as long as the distance between true and initial landmark positions is well within $kn_s 2^{(N_{max}-1)}$ pixels. But if the object can be located anywhere within the input image, an (exhaustive) search to find a suitable initialization, *e.g.* as described in [25], may be necessary.
7. If information is available about the confidence of the proposed landmark displacement, weighted fitting of the shape model can be used, as explained in [47].
8. Standard non-linear optimization algorithms, such as hill climbing, Levenberg-Marquardt, or genetic algorithms can be used to find the optimal model parameters \mathbf{b} instead of using the alternating displacement of landmarks and model fitting as proposed by Cootes and Taylor. Note that a multi-resolution approach can still be used in such a framework. Alternatively, a snake algorithm can be used as well in which the shape model provides an internal energy term and the appearance model fit is used as external energy term. The use of different optimization algorithms is not considered in this chapter.
9. Behiels *et al.* [17] propose to use dynamic programming to find new positions for the landmarks after the Mahalanobis distance at each new possible position has been computed. This avoids outliers among the displaced landmarks and leads, according to [17], to quicker convergence.

ASMs with optimal features

In this section a new appearance model is described that is an alternative to the construction of normalized first derivative profiles and the Mahalanobis distance cost function of the original ASMs.

The aim of the appearance model is to be able to move the landmark points to better locations during optimization. The best location is the one for which all

points on one side are outside the object, and those on the other side are inside of it. Therefore the probability that a point is inside/outside the object is estimated, for each point separately. We base this classification on *optimal* local image features, obtained from feature selection instead of the fixed choice of the normalized first derivative profile, using a k NN-classifier.

Let us first focus on the features. We are looking for general image structure descriptors. A Taylor expansion approximates a function f around a point of interest x_0 by a polynomial of (some) order N . The coefficients in front of each term are given by the derivatives $f^{(n)}$ at x_0

$$f(x) \approx \sum_{n=0}^N \frac{1}{n!} f^{(n)}(x_0) (x - x_0)^n. \quad (6.8)$$

Derivatives of images are computed by convolution of derivatives of Gaussians. This motivates the use of a filter bank of Gaussian derivatives to locally describe image structure. The derivatives can be measured at various scales. The Taylor expansion of images is known as the local jet, or multi-scale local jet in the case of the Taylor expansion of the scale-space of the image [137, 70].

We extract local texture features from the derived images by computing moments of local histograms. When a histogram is computed, three scale parameters are important. The first is the inner scale σ at which the image is observed; this parameter is already included in the computation of the derivative. The second is the scale of the histogram, the tonal scale β . The third is the extent α over which the histogram is computed. Locally orderless images, as developed by Koenderink and Van Doorn [141] and described in Chapters 4 and 5, provide a formal definition of local histograms that is consistent with scale-space requirements.

The key idea to locally orderless images is to replace the image L which gives the intensity i at each location \mathbf{x} with a family of *soft isophote* images which give the probability of the occurrence of an intensity i at \mathbf{x} :

$$R(\mathbf{x}, i; \sigma, \beta) = \exp\left(-\frac{(L(\mathbf{x}; \sigma) - i)^2}{2\beta^2}\right). \quad (6.9)$$

If a number of such images R are computed and blurred to a scale α then we have at each \mathbf{x} a local histogram with (Gaussian!) extent α . Moments m can be computed for each location. Notice that there are quite some parameters to vary (m , the order of derivatives in the x - and y -direction, σ , α , β) and it remains an open question which combinations are optimal for a given application and even a given location in the images. Our strategy is to compute an extensive set of features and use feature selection techniques in the subsequent classification stage to determine the optimal features. However, some combinations are generally not sensible. The first moments of derivatives ($n > 0$) are close to zero and do not contain textural information. Another constraint is that we should have $\alpha > \sigma$ otherwise the histogram will not be interesting. It can also be shown that if only moments of histograms are used, no

new information is extracted by using multiple scales for β . See Chapter 5 for more details.

In this work we use only first and second moments ($m = 1, 2$), all derivatives up to second order ($L, L_x, L_y, L_{xx}, L_{yy}, L_{xy}$), five inner scales ($\sigma = 0.5, 1, 2, 4, 8$), and a fixed relation between the inner scale σ and the histogram extent α of $\alpha = 2\sigma$. For the first moments this yields an effective scale of 1.12, 2.23, 4.47, 8.94, and 17.89, respectively. The total number of features is $2 \times 6 \times 5 = 60$.

Obviously the method can be extended by using more scales and higher order derivatives, higher moments, or releasing the fixed relation between σ and α .

The next step is to specify how to construct a training set from the training images, which classifier to use, and how to do feature selection.

Consider again the optimization procedure. At each iteration, each landmark is positioned at $2k + 1$ locations along a profile perpendicular to the current object location. Obviously the image structure is different for each landmark, but the positions that are evaluated are also different for each resolution. Therefore we will select an optimal set of features for each landmark *and* each resolution separately. Note that this is the same in the original ASMs: the covariance matrices \mathbf{S}_g as they appear in Eq. (6.7) are computed for each landmark and each resolution.

From each training image and for each landmark a square grid of $N_{grid} \times N_{grid}$ points is defined with N_{grid} an odd integer and the landmark point in the center of the grid. The spacing is $2^{(i-1)}$ pixels for the i th resolution level.

In this work we use $N_{grid} = 5$ which means that for each landmark and each resolution level, a feature vector with 60 elements is sampled at 25 points. The output of each feature vector is either inside (1) or outside (0) the object. The landmark points itself are considered to be inside the objects (this is a rather arbitrary choice). The set of training images is divided in two equal parts. This leads to two sets of samples, a training and a validation set. A k NN classifier with weighted voting is used. $k_{NN} = 5$ was used and the weight of each vote is $\exp(-d^2)$, where d is the distance to each neighbor.

Sequential feature forward selection (SFFS) (also known as Whitney's method, see [264, 203, 223] and Chapter 5) is used to find an optimal feature set of at most f_{max} features. Sequential feature backward selection is used on the optimal feature set to see whether eliminating features can further improve performance. After feature selection, the samples from the training and the validation set are merged and a list of selected features for each landmark and each resolution is stored.

When the model is fitted to an input image, the scheme starts with computing the 60 feature images; an example of the result is given in Figure 6.1. Instead of sampling the normalized derivative profiles, the optimal feature set at each position along the profile is fed into a k NN classifier to determine the probability that this pixel is inside the object. The objective function to be minimized is the sum of absolute differences between the expected probability (0 or 1 for points outside or inside the object, respectively) and the predicted probability, for each point along the profile. This is illustrated in Figure 6.2.

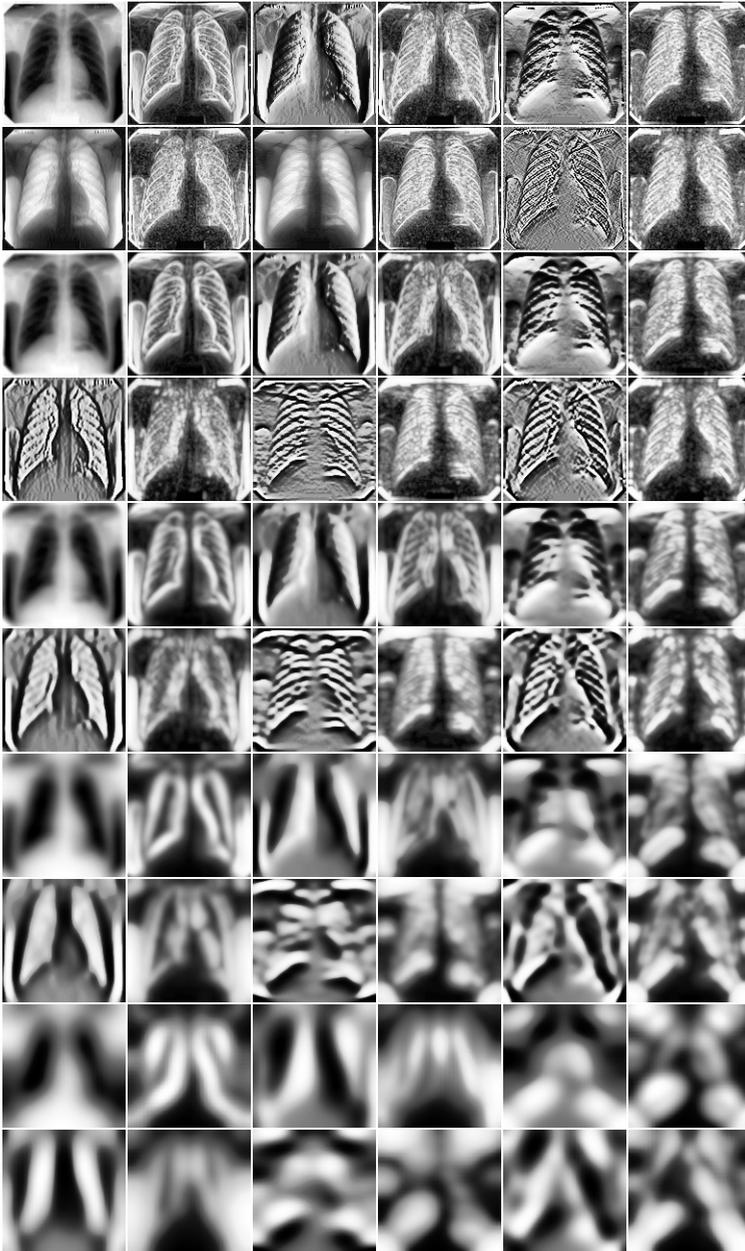


Figure 6.1: Illustration of the 60 feature images for one input image, shown top-left. The first two rows show the results for $\sigma = 0.5; \alpha = 1$. The first row shows the first and second moments of L , L_x , and L_y . The second row shows the first and second moments of L_{xx} , L_{yy} , and L_{xy} , respectively. Subsequent rows display the results for $\sigma = 1; \alpha = 2$, $\sigma = 2; \alpha = 4$, $\sigma = 4; \alpha = 8$, and $\sigma = 8; \alpha = 16$, respectively.

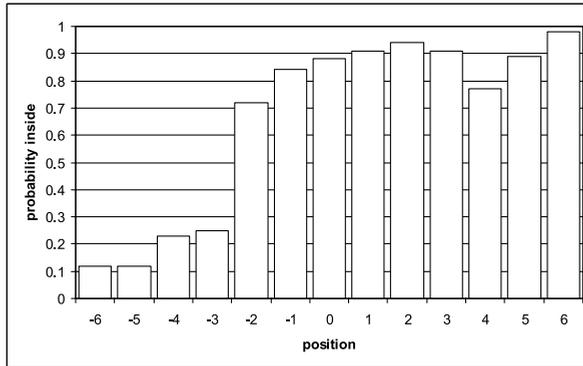


Figure 6.2: Illustration of the procedure used to displace landmark points in the ASM scheme with optimal features. The probability that a position is inside the object is evaluated at a number of points on the line perpendicular to a landmark. In this example 9 new positions of the landmark need to be evaluated ($n_s = 4$) and the profile length is 5 ($k = 2$). The landmark is originally located at position 0 and the probability of $n_s + k = 6$ positions to the left and right are computed. The graph shows the probabilities. Ideally, the landmark should be placed at a position with profile 0, 0, 1, 1, 1 (for a profile of length 5). All positions from -4 to +4 are evaluated and the new optimal position for the landmark is at position -2.

Computational considerations

One of the advantages of the original ASM scheme compared to other segmentation methods is its speed. The new method requires considerably more computation time. However, all the feature selection can be done off-line. An optimized k NN classifier [8] was used, available at <http://www.cs.umd.edu/~mount/ANN>. The feature images have to be computed on-line, which required 8.0 seconds on a 600 MHz Pentium PC for a 256 by 256 image. For the original ASM method, a number of blurred images have to be computed, which required 0.35 seconds on a 600 MHz Pentium PC for a 256 by 256 image.

During optimization feature vectors must be classified by k NN classifiers and this requires more time than computing Mahalanobis distances. The total time for the segmentation of 256 by 256 resolution images was 0.17 seconds for the original ASM scheme and 4.1 seconds for the method with optimal features. All these benchmark figures have been obtained with the same parameter settings that are used in all experiments in this chapter.

Using a smaller feature set would reduce the computational cost of the method (almost linearly). An alternative would be to select a subset of the 60 features for all operations (each landmark, each resolution) so that it is no longer necessary to compute all 60 images for each input image. However, this speed improvement would probably come at the price of a small decrease in performance.

Experiments

Materials

Experiments have been carried out with three types of images and five different objects.

The first database consisted of 80 simulated images with an object that has the shape of a house from which the five corner points were perturbed. The house and the background were filled with different textures. Experiments on these simulated data were included to demonstrate that the method with optimal features can deal with such textural boundaries while the original method cannot. Details about the construction of the images are given in Figure 6.3.

The second database consisted of 230 standard PA chest radiographs. In these images both the left and right lung fields were segmented, as separate objects. The images were randomly selected from a tuberculosis screening program and contained both normal and abnormal cases of patients of 16 years and older. The images were taken with a mobile Oldelft Electrodelca (Nucletron BV, Veenendaal, The Netherlands). The tube voltage was 117 kV and the images were printed on 10 by 10 cm film. The films were digitized with a Lumisys 100 scanner (Lumisys, Inc., Sunnyvale, CA) to 996 by 996 pixels with 12 bit intensity, subsampled to 256×256 pixels.

The third database was a collection of 90 MRI slices of the brain, in which the corpus callosum and the cerebellum were segmented. The images were made available by the University of Iowa Hospitals and Clinics and were also used by Brejl and Sonka in [25]. The resolution is 320×256 pixels, 0.7 mm per pixel, obtained by interpolating the original volumetric data acquired with 1.5 mm slice thickness and 0.7 mm^2 coronal plane resolution.

The objects in the images were annotated by a number of *fixed* landmarks and

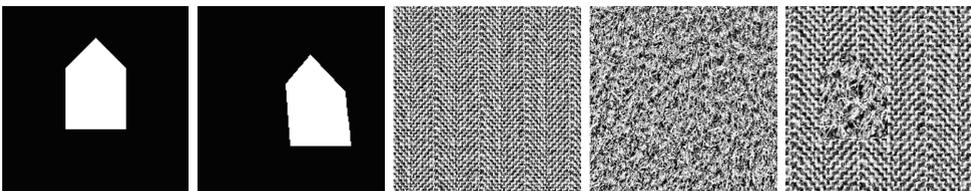


Figure 6.3: (a) A generic house image. The image has a resolution of 200×200 pixels. New house images are randomly generated by adding a displacement (dx, dy) to each corner point where dx and dy are randomly selected from the interval $(-20, +20)$ pixels, and subsequently translating the distorted house by (dx, dy) where dx and dy are randomly selected from the interval $(-35, +35)$ pixels. (b) An example of a generated house image. (c) The outside of the house is filled with a part of this texture image (resolution 300×300 pixels) using a random offset. (d) The inside of the house is filled with a part of this texture image (resolution 300×300 pixels) using a random offset. Both texture images are taken from the Brodatz set [26] and histogram equalized. (e) An example of a generated textured house image.

a closed contour between those fixed points from which a number of equidistant landmark points were sampled. For the house images, 5 fixed landmark points and 26 points in total were used. For the lung fields, the number of fixed points was 3 and in total 40 points were used. The corpus callosum was indicated with 3 fixed landmarks and 50 points in total. The cerebellum was given by 7 fixed landmarks and 50 points in total.

Methods

For each parameter of ASMs, a fixed setting was selected. For the house and lung shapes, no shape alignment was performed (this improved performance) and a shape model was constructed in which 99.5% of the variance was explained ($f_v = 0.995$). For the brain structures, shape alignment was used (in this case better results were obtained with the use of alignment) and a shape model explaining 98% of the variance ($f_v = 0.98$) was constructed. The fact that f_v is lower for the aligned brain shapes is because alignment reduces the amount of variation between shapes and a higher proportion of the data can be attributed to noise.

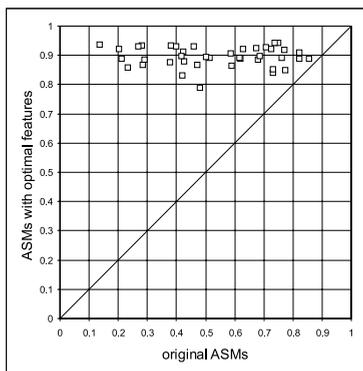
The other settings were 4 levels of resolution ($L_{max} = 4$), 10 iterations per level ($N_{max} = 10$), profiles of length 5 ($k = 2$) and evaluation of 9 positions per iteration ($n_s = 4$). When fitting the shape model to the displaced landmarks, each mode was constrained within 2 times the standard deviation ($m = 2.0$). For the extended ASMs, at most 10 features were selected for each landmark and each resolution ($f_{max} = 10$). Training data were selected from 5 by 5 neighborhoods around each landmark ($N_{grid} = 5$). In the k NN classifier, 5 neighbors were used ($k_{NN} = 5$). All parameter settings are also given in Table 6.1.

To evaluate the results, the following “overlap” measure was used: the area correctly classified as object was divided by the area classified as object plus the area incorrectly classified as background. In other words: true positive area divided by true positive plus false positive plus false negative area. This performance measure yields 1 for a perfect result and 0 if there is no overlap at all between the detected and true object. This measure is probably more directly related to what is expected of a segmentation in practice than the average distance between the true and detected location of each landmark in the shape model, because the latter is not sensitive to shifts of the landmarks along the contour.

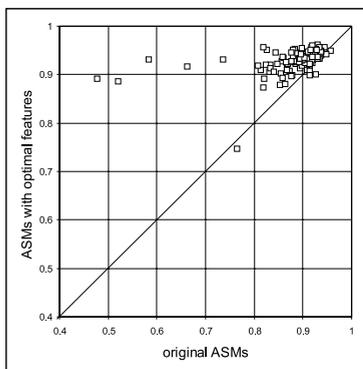
In all experiments the performance when fitting the shape directly to the true landmarks (*cf.* Eq. (6.5)) was also computed. The lung fields in the test set of the chest radiographs were manually segmented by a second observer independently. Therefore the manual segmentation by the second observer could be compared to the output of the automatic methods. We refer to this comparison between the two observers as the inter-observer variability.

Results

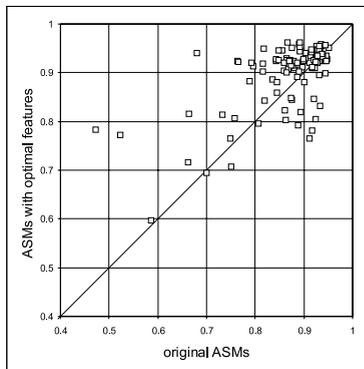
The results of all experiments are summarized in Table 6.2. The result of directly fitting the shape model to the landmark points is included because it indicates an



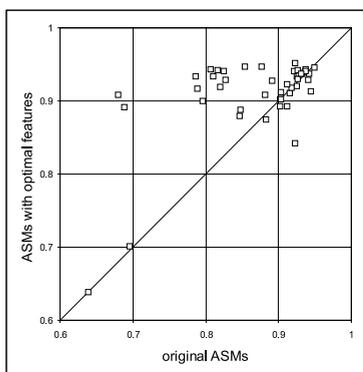
Houses



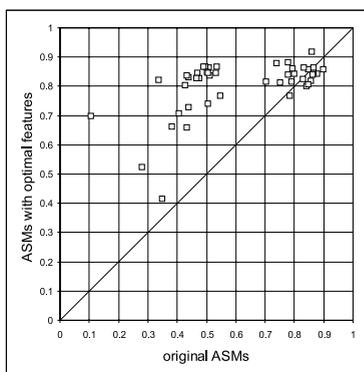
Right lung fields



Left lung fields



Cerebellum



Corpus callosum

Figure 6.4: Scatterplots of each segmentation experiment. The overlap measure for the original ASM scheme is plotted against the measure for the ASM method with optimal features for each segmented image.

Houses	$\mu \pm \sigma$	median
ASMs	0.536 ± 0.203	0.550
ASMs with optimal features	0.895 ± 0.034	0.894
Fit of shape model	0.933 ± 0.029	0.937
Right lung	$\mu \pm \sigma$	median
ASMs	0.882 ± 0.074	0.902
ASMs with optimal features	0.929 ± 0.026	0.933
Fit of shape model	0.948 ± 0.030	0.955
Second observer	0.945 ± 0.017	0.948
Left lung	$\mu \pm \sigma$	median
ASMs	0.861 ± 0.109	0.891
ASMs with optimal features	0.887 ± 0.114	0.924
Fit of shape model	0.942 ± 0.090	0.955
Second observer	0.934 ± 0.021	0.938
Corpus callosum	$\mu \pm \sigma$	median
ASMs	0.617 ± 0.206	0.535
ASMs with optimal features	0.805 ± 0.093	0.837
Fit of shape model	0.887 ± 0.052	0.906
Cerebellum	$\mu \pm \sigma$	median
ASMs	0.870 ± 0.078	0.904
ASMs with optimal features	0.910 ± 0.058	0.927
Fit of shape model	0.950 ± 0.014	0.950

Table 6.2: Mean, standard deviation and median results of all experiments. The original ASM scheme is compared with the method with optimal features and the result of directly fitting the shape model to the true landmark positions. In the case of the chest radiographs the mean, standard deviation and median results of a second human operator are also given. All error measures are expressed as the ratio between the true positive area and the true positive area plus false positive area plus false negative area.

upper bound for both the original method and the method with optimal features. Note that fitting the shape model minimizes the distance between the predicted landmark position and the true landmark position; it does not necessarily optimize the overlap measure. Therefore it is possible that an ASM scheme produces a set of model parameters \mathbf{b} for which the overlap measure is higher than the one found by fitting the shape model according to Eq. (6.5). This occurred in a few cases. Another practical measure of the optimal performance any automatic segmentation method that is trained with examples can achieve is the variation between observers. This measure is given for the lung field segmentations. For this application, the median result of the ASM method with optimal features is close to median result of a second

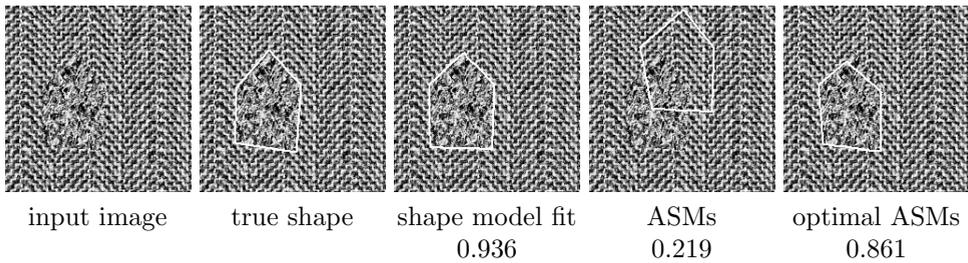


Figure 6.5: An example result for the segmentation of the generated textured images with a house shape. Segmentations are given by the thick white line. Below each segmentation the overlap measure is given. The original scheme clearly cannot deal with such images. ASMs with optimal features produces a reasonable result.

human observer.

Table 6.2 shows that the original ASM scheme is not able to segment the simulated house images, contrary to the scheme with optimal features. For the other datasets, the differences between the methods are smaller, but in all cases ASMs with optimal features produced significantly better segmentations than the original scheme ($p < 0.001$ in a paired T-test for all experiments). This is also clear from Figure 6.4 which shows scatterplots for each segmentation task. In these plots, points which are above the diagonal line indicate images for which the segmentation with optimal features is better than the result of the original scheme. Again it is evident that the biggest differences between both methods occur for the house images. In all other experiments the general improvement of using optimal features method is clear, but in case of the left lung fields there is also a considerable number of cases where the original method has better performance.

A typical result for the house image segmentation is shown in Figure 6.5. Figures 6.6 and 6.7 show typical results for the right lung field and left lung field segmentation.

For the MR brain images two example results for both the corpus callosum and cerebellum segmentation are shown in Figures 6.8 and 6.9. In the middle rows cases are shown where the original scheme seriously failed but ASMs with optimal features produced reasonable results. More representative are the cases in the bottom rows of Figures 6.8 and 6.9 in which the optimized scheme produces only slightly better results.

Discussion and conclusions

Active shape models have been applied to several segmentation tasks in medical imaging, mainly on X-ray images and slices from MR and CT data. The method has been shown to produce good results in many cases. This is also the case for the four medical segmentation tasks considered in this chapter. The main result of this work is that the results can be improved by replacing the border appearance model in the original method with a non-linear classifier and a selection mechanism that finds optimal local

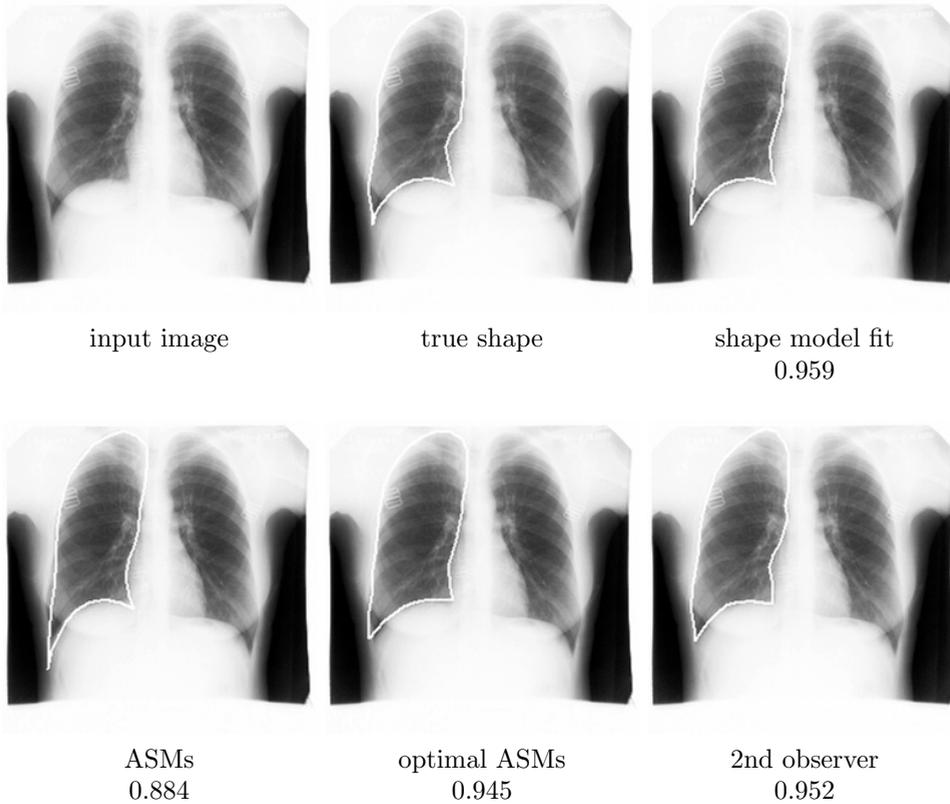


Figure 6.6: A typical example result for the right lung field segmentation. Segmentations are given by the thick white line. Below each segmentation the overlap measure is given. The scheme with optimal features produces slightly better results than the original method. Fitting the shape model directly to the true landmarks yields an overlap measure comparable to the result of a 2nd observer.

features from the training data.

The biggest performance improvement is obtained on simulated data in which the textural appearance of the image inside and outside the object is different. This indicates that the proposed method may be especially useful to segment textured objects from textured backgrounds. An example could be segmentation in ultrasound images.

An important aspect is that an optimal set of features is selected for each landmark and each resolution separately. Alternatively, a single set of optimal features could be used, which would be equivalent to designing a pixel classifier that assigns each image location to one of two classes: outside or inside the object of interest. The segmentation method could be run on these “potential images” instead of on the real data. Pilot experiments indicated that the performance of such an approach is lower than that of the method presented here, ASMs with optimal features. The set of

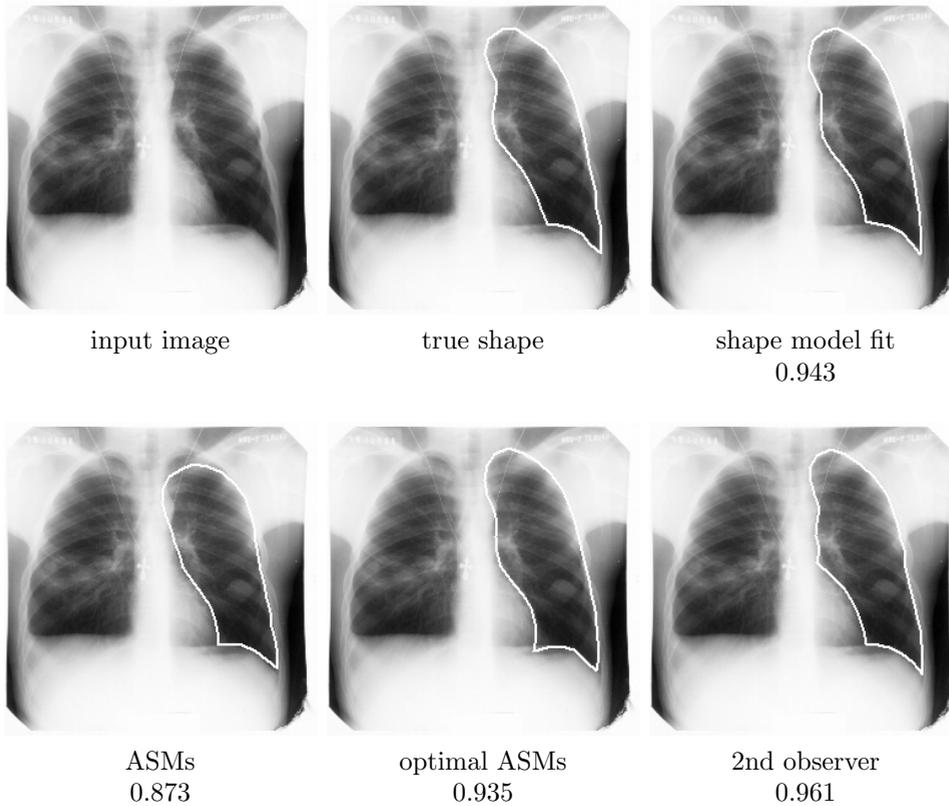


Figure 6.7: An example result for the left lung field segmentation. Segmentations are given by the thick white line. Below each segmentation the overlap measure is given. The original scheme incorrectly detects the lung top and does not produce a correct bend around the aortic arch. ASMs with optimal features segments the left lung field quite well. Note that the shape of the right lung field is abnormal and that both lung fields contain abnormalities.

selected features varies considerably per landmark and resolution and is different for different applications.

There are many parameters in the original scheme, and the new approach even introduces a few more. We have not thoroughly investigated the effect of varying these parameters and therefore it is not known whether the used settings are optimal, although small changes in the parameters usually did not influence the results very much. In practice it could be desirable to perform some sort of (semi-)automatic selection of optimal parameters on the training set.

The ASM method with optimal features performed significantly better in all applications. Although the improved method produces results in a few seconds, it requires a 20-fold longer computation time than the original ASMs. This can be reduced if the feature images are computed only at those points where their value is required

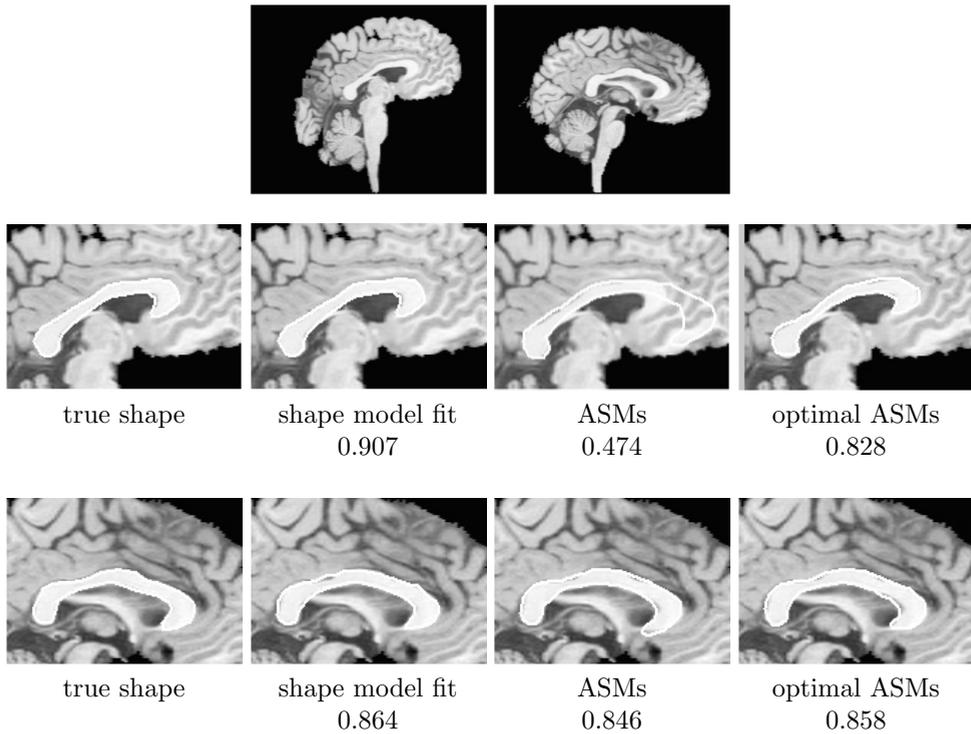


Figure 6.8: Two results of segmenting the corpus callosum. The top row shows the two input images. The results for the left and right image are shown in the middle and bottom row, respectively, with the segmentation results as a thick white line. From left to right the true (manual) segmentation is shown, the result of directly fitting the shape model to the true landmark positions, the result of the original scheme and the result of ASMs with optimal features. The overlap measure is indicated below each image. The middle row is an example of a case where the regular method does not converge to a correct solution because the right part of the shape was attracted to an incorrect part of the image during the first few iterations of the scheme, at a coarse resolution. The bottom row shows a case in which both methods produce a reasonable segmentation.

during segmentation. Using a faster classifier will also reduce computation time.

The ASM scheme with optimized features could be improved as follows (where it should be noted that the list of improvements is by no means exhaustive):

- More feature images can be used, by using higher moments, more (higher-order) derivatives and by relaxing the fixed relation between σ and α .
- The landmarks will be moved to those locations where the profile values are closest to 0 for points outside the object and closest to 1 for points inside the object. In practice the optimal profiles may be different. Especially if

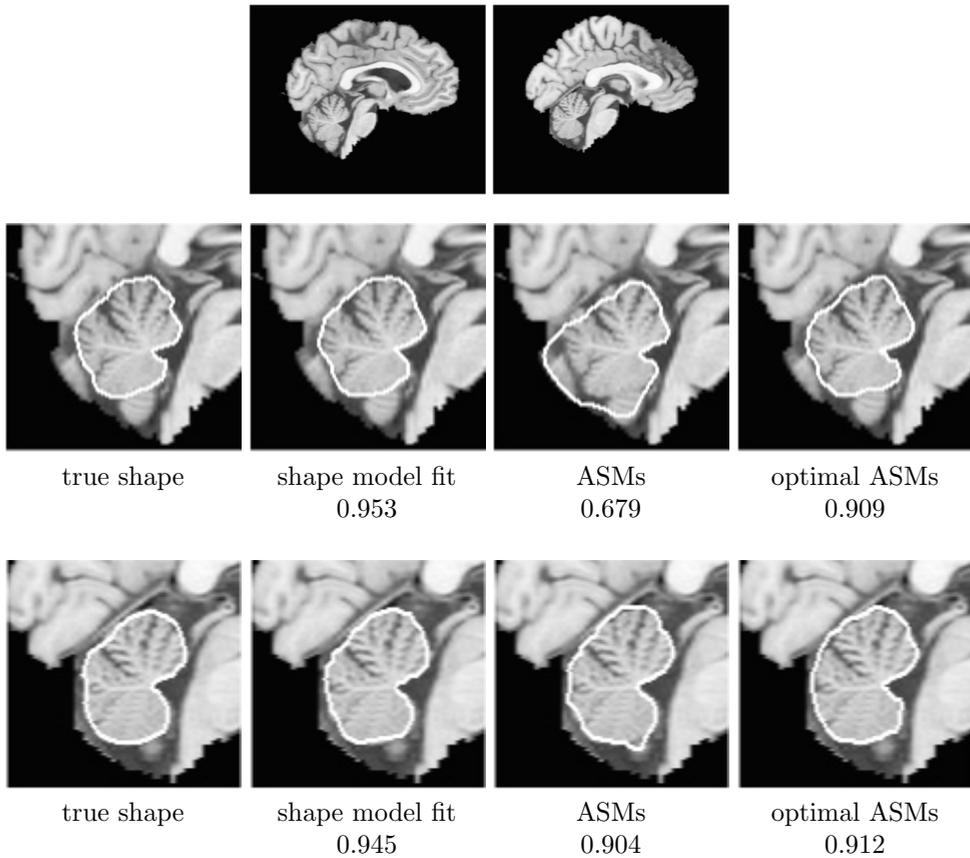


Figure 6.9: Two results of segmenting the cerebellum. The top row shows the two input images. The results for the left and right image are shown in the middle and bottom row, respectively, with the segmentation results as a thick white line. From left to right the true (manual) segmentation is shown, the result of directly fitting the shape model to the true landmark positions, the result of the original scheme and the result of ASMs with optimal features. The overlap measure is indicated below each image. The middle row is an example of a case where the scheme with optimal features produces a good result but the regular method includes two areas on the left and below the cerebellum in its segmentation. The bottom row shows a case in which both methods produce comparable results.

the object is very thin and the fitting occurs at a coarse resolution level, the innermost points of the profile may cross the border on the other side of the object. A straightforward extension would be to extract the actual profiles from the training set and construct covariance matrices and measure the suitability of a profile with the Mahalanobis distance, as in the original ASM scheme.

- The current scheme does not take the structure along the profile into account,

but uses local pixel classification for each position along the profile independently. Performance may increase in some cases if the feature set at *all* positions along the profile is used directly to classify the position of the landmark. Consider a set of images, half show a black object on a white background, and the other half a white object on a black background. With local image features, it is impossible to classify locations as inside or outside the object. But a set of features along the profile can easily distinguish correct profiles, with an intensity jump at exactly the landmark location, from incorrect profiles, with no intensity jump or an intensity jump at a different location.

- Dynamic programming can be used to find the optimal displacements for landmark positions [17] instead of simply taking the optimal position for each landmark independently.
- Similarly, a projection of the shape model that minimizes only the distances between landmark points *perpendicular* to the contour [105] can be used instead of Eq. (6.5).

Chapter 7

Segmenting the Rib Cage

An automatic method for segmenting the posterior ribs in frontal chest radiographs is presented, based on a statistical shape model for the complete rib cage. Contrary to previous work, we fit the global rib cage directly to a radiograph, instead of detecting rib border candidates locally and applying rules to infer the rib cage from these candidates. Each posterior rib is modeled by two equidistant parabolas. The full rib cage, from rib 2 up and including rib 10, therefore contains 72 parameters. This number is reduced with principal component analysis; it is demonstrated that 10 parameters explain over 98% of the variability in a training set of 35 chest radiographs. The rib cage is fitted with Powell's direction set method for optimizing the model parameters, with a fit measure that accumulates the edge strength perpendicular to the rib borders at a fixed scale. The method is tested on 35 images with a resolution of 512 by 512 pixels. The result is that rib borders are located with an accuracy of 3 pixels on average.

“THE RIBS CONFUSE EVERYONE beginning to look at chest films. The miracle is that one can discern anything useful about the heart and the lungs through such a crosshatched pattern of shadow.” This is a quote from a textbook on radiology [225] that not only applies to students of radiology, but to computer vision algorithms as well. There are several reasons why delineating the ribs in chest radiographs is not an easy problem. First of all, a projection image of the rib cage produces a pattern of overlaid structures including crossings of posterior and anterior ribs. Second, there exists considerable anatomical variation between subjects, and their habitus during the examination determines the way in which the ribs are projected. Furthermore, standard chest radiographs are usually obtained with a soft beam that does not produce a clear outline of the ribs, but is optimized for visualizing softer tissues.

Not only this complexity makes automatic segmentation of ribs in thorax images an interesting problem; it is also a task with great practical importance. Chest radiographs remain the initial and often most valuable means of chest examination, making up 40% of all x-ray examinations in total. The ribs provide a convenient

frame of reference in the chest. Radiologists frequently denote locations in the chest by referring to one of the costal or intercostal spaces. Almost any computer analysis of chest images will therefore benefit from an accurate delineation of the rib cage. And this delineation may also help to remove false positives that are due to (possibly crossing) rib structures in computerized detection of abnormalities in chest images.

Relation to previous work

This section briefly reviews previous work on rib segmentation. A more detailed overview is given in Chapter 2.

Attempts to automatically detect ribs in frontal chest radiographs go back as far as 1973 when Toriwaki *et al.* [233] used template matching to find likely locations of rib borders and fitted parabolas to these locations. Wechsler [256, 257] used edge detection and thresholding to find rib border candidates and proposed a slightly more elaborate rib model, consisting of a [55] used the derivatives of vertical profiles in the image to find pixels on rib boundaries and fitted parabolas to these. A general problem emerged from all these studies: not all detected edge pixels and extracted rib border candidates were actually located on a true rib border. In order to distinguish real rib borders from other structures in the image, several heuristic reasoning schemes have been proposed. Ballard [14] used a relational graph to represent the relationships between geometrical primitives in the image and to infer the location of more obscure ribs from ribs that stand out more clearly. Powell [201], Sanada [211] and co-workers incorporated a model of the full rib cage by fitting a sinusoidal function with spatially varying frequency to vertical profiles of chest images. The varying frequency was supposed to account for the varying spacing between ribs.

Broadly speaking, more recent approaches follow the same strategies. Yue *et al.* [278] fitted parabolas to edge pixels with a Hough transform, used simple rules to discriminate “true” maxima in the Hough space from “false” responses and refined the location of the rib borders with snakes. Vogelsang *et al.* [253] compared the method of fitting parabolas to edges using a Hough transform with a template matching approach, again followed by simple rules to determine which parabolas correspond to ribs, using estimates of the parameters of the parabolas.

The method presented here contains several of the classical ingredients of all these previous studies, but differs in an important way: we derive a statistical model of the complete rib cage and fit it directly to the image. The statistical model is obtained with principal component analysis. This is a global approach that has no need for the ad hoc reasoning stage of earlier work.

Problem definition

The aim is to automatically segment the posterior rib cage from a regular frontal chest image. See Figure 7.1 for an illustration of the rib cage and the distinction between posterior and anterior ribs. The first rib is ignored, because it is visible as an anterior rib, and the eleventh and twelfth rib are ignored because they are usually not visible on a chest radiograph. So the posterior parts of all ribs from the second

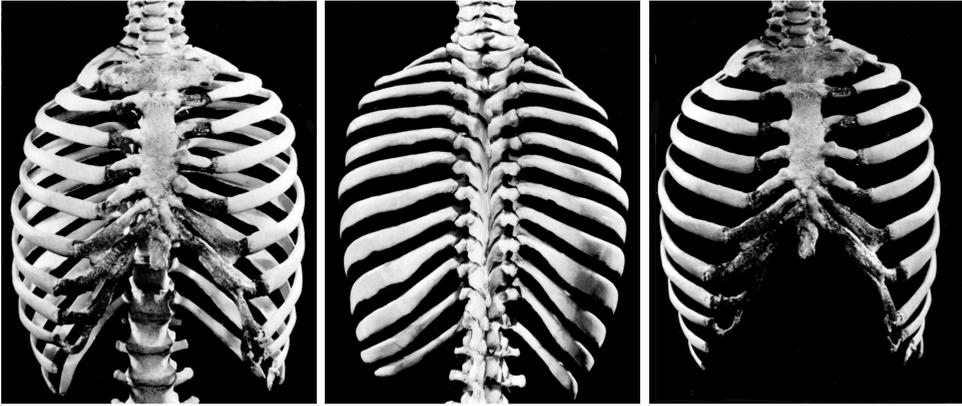


Figure 7.1: Photographs of the bony thorax. Left: the complete rib cage, photographed from the front side. Middle: only the posterior or dorsal ribs, photographed from the back. The thoracic cage was stuffed with velvet to produce this photo. Right: only the anterior or ventral part of the rib cage (again with the thoracic cage stuffed with velvet). From these photos, it is easy to appreciate that the ribs that are most visible in a chest radiographs are the posterior ribs. Photos are reproduced from [225].

through the tenth are detected. The locations of the lung fields are supposed to be known. Although this may seem to be a restriction of the applicability of the method, in practice a computerized analysis of a chest image will usually segment the lung fields before attempting to delineate the ribs. In this paper, we used the lung fields segmented by the method described in Chapter 3 and in [82, 84].

Materials

Thirty-five standard PA chest radiographs are taken from a tuberculosis screening program for people seeking political asylum in The Netherlands. The images contain both normal and abnormal cases and were taken with a mobile Oldelft Electrodelca (Nucletron BV, Veenendaal, The Netherlands), a system commonly used in mass chest screening. The tube voltage was 117 kV and the images were printed on 10 by 10 cm film. The films were digitized with a Lumisys 100 scanner (Lumisys, Inc., Sunnyvale, CA) to 996 by 996 pixels with 12 bit intensity, and subsampled to 512 by 512 pixels.

Rib model

Each rib border is modeled as a parabola. The upper and lower border are vertically shifted versions of the same parabola. We refer to this as equidistant parabolas. The parameterization is as follows. The center of mass of each lung field is determined.

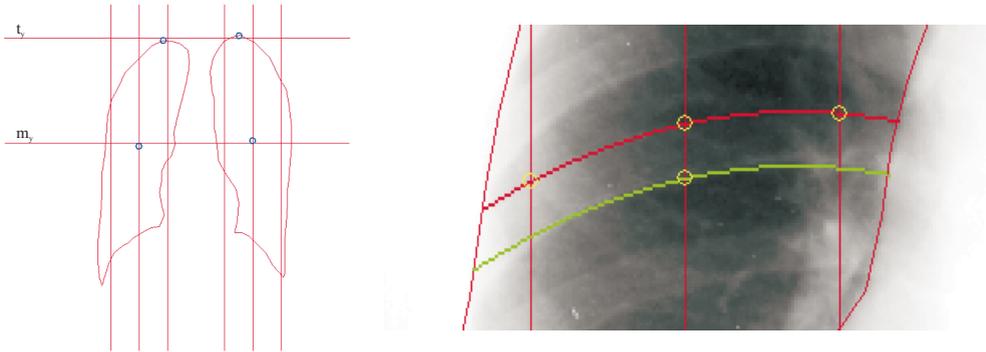


Figure 7.2: Left: the reference frame with six landmarks lines, as computed from the centers of mass and tops of the lung fields. See the text for an explanation. Right: each rib is modeled as two equidistant parabolas, defined by four points on the landmark lines. Three points – one on each landmark line – define the upper rib boundary. A fourth point on the middle landmark line defines the lower border.

The heights of these centers are averaged and given by m_y . The heights of the lung tops – defined as the highest point of each lung field – are also averaged and given by t_y . All y -coordinates are scaled so that they are 0 for m_y and 1 for t_y . We define 6 vertical landmark lines. Two vertical lines are drawn through the centers of mass of both lung fields. The horizontal distance between the centers of mass of both lung fields is divided by 4, and on both sides of both landmark lines, another landmark line is drawn at this distance. Figure 7.2 illustrates the six landmark lines and the scaled y -coordinate system. Each rib is now represented by the scaled y -coordinate of 4 points on these landmark lines. Three points define a parabola for the upper rib border; the fourth point is located on the lower rib border and thus defines the width of the rib. This is also illustrated in Figure 7.2. The full rib cage is represented by 9 ribs on both sides, with 4 points per rib, so a total of 72 parameters.

The advantage of parameterizing the ribs relative to the lung fields in this manner is that the parameters defining each rib vary relatively little from image to image. This will turn out to be useful in the subsequent modeling of the rib cage.

Statistical rib cage model

In order to reduce the number of degrees of freedom in the model, principal component analysis [121] is performed on the 72 parameters that define the rib cage. This technique has recently found widespread application in computer vision, where landmark points are used to describe shapes and extracted active shape models are fitted to images [63, 49, 48]. Note however that in this case the parameters are not landmark points: they are merely the (scaled) y -coordinates that define sets of equidistant parabolas.

The parameters $p_1 \dots p_N$ (where $N = 72$ in the case of our rib cage model) are

stacked into a vector \mathbf{p} . The vector for the i th training image is denoted with \mathbf{p}_i . The mean parameter vector is computed

$$\bar{\mathbf{p}} = \frac{1}{s} \sum_{i=1}^s \mathbf{p}_i, \quad (7.1)$$

and the covariance matrix

$$\mathbf{S} = \frac{1}{s-1} \sum_{i=1}^s (\mathbf{p}_i - \bar{\mathbf{p}})(\mathbf{p}_i - \bar{\mathbf{p}})^T. \quad (7.2)$$

The eigenvectors ϕ_i of this covariance matrix, together with their corresponding eigenvalues λ_i are sorted so that $\lambda_i \geq \lambda_{i+1}$. If the t largest eigenvalues and their eigenvectors are retained to form the matrix $\Phi = (\phi_1|\phi_2|\dots|\phi_t)$, any parameter vector of the training set can be approximated with the projection

$$\mathbf{p} \approx \bar{\mathbf{p}} + \Phi \mathbf{b}, \quad (7.3)$$

where the vector \mathbf{b} contains the projections of \mathbf{p} along the t principal components of the cloud of points $\mathbf{p}_1 \dots \mathbf{p}_s$ in ND space. The model parameters \mathbf{b} can be computed from

$$\mathbf{b} = \Phi^T(\mathbf{p} - \bar{\mathbf{p}}). \quad (7.4)$$

The total variance in the training data is given by the sum of the eigenvalues $\sum_{i=1}^s \lambda_i$. Thus by retaining t modes $\sum_{i=1}^t \lambda_i / \sum_{i=1}^s \lambda_i$ of the total variance in the training set is explained.

The eigenvalues λ_i denote the amount of variation in the training set along each direction ϕ_i . Therefore we can generate plausible model instances with the projection formula (7.3) by applying limits $\pm 3\sqrt{\lambda_i}$ on each parameter b_i in the model.

The author manually indicated the 4 points on the landmark lines that define the ribs of each of the 35 images in the training set. PCA was performed on the vectors and in Figure 7.3 the explained variance as a function of the number of retained modes is plotted, together with the eigenvalues. With 10 modes over 98% of the total variance can be explained.

The statistical model of the rib cage is interesting in itself. Figure 7.4 shows the rib cage as generated by varying the first 4 modes of the model from $-3\sqrt{\lambda_i}$ (left column) to 0 (middle column) to $+3\sqrt{\lambda_i}$ (right column). The modes can be interpreted directly. The first mode scales the height of the rib cage relative to the lung fields (since the coordinates of the ribs are expressed relative to the height of the lung fields). Not surprisingly, this is the most significant mode of variation among the training images. The second mode reveals that in some images the space between the lower ribs is larger than between the first few ribs, while in other cases the ribs

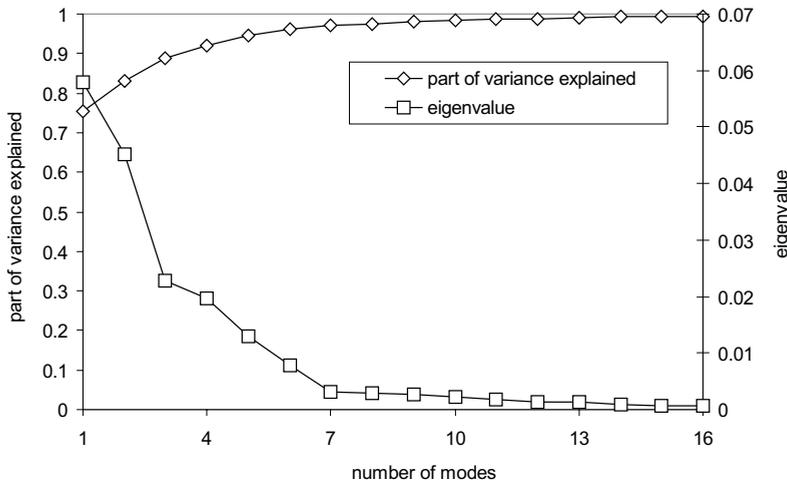


Figure 7.3: Principal component analysis is used to obtain a statistical model of the rib cage. By adding more modes to the model, more variance is explained. The graph shows the amount of variance explained as a function of the number of modes. The corresponding sorted eigenvalues of the covariance matrix are also plotted. The total number of modes needed to explain over 98% of the variance is 10.

are spaced more equally than in the mean rib cage. Moreover, this mode also controls the shape of individual ribs, from highly curved to almost straight. The third mode is related to asymmetry between the right and left part of the rib cage. This can be the result of the habitus of the patient during the examination. The fourth mode shows that the thickness of the ribs relative to the inter-rib spacing varies from subject to subject, and that the third rib is sometimes shaped like the second rib and in other cases more like the fourth rib.

Given the manual segmentation of a rib cage (the true rib cage) and a rib cage generated from the statistical model, we define the following error measure: for each of the four points per rib in the true rib cage, find the closest corresponding point of the generated rib cage, and sum the absolute distances (in the scaled y -coordinate system). The reason for finding the closest point instead using the corresponding point is that the system may skip a rib during fitting after which corresponding points of all subsequent ribs are located on different ribs and the error is very large, although the rib borders may coincide very well.

Figure 7.5 shows this error measure when each of the rib cages in the training set is approximated by the model, using the projection formula (7.3), as a function of the number of modes in the model. To do this, the model was recomputed for all shapes in the training set with the rib cage to fit left out. This leave-one-out strategy has been used in all other experiments presented in this work. Figure 7.6 shows the result for a single image for several numbers of retained modes, fitting the

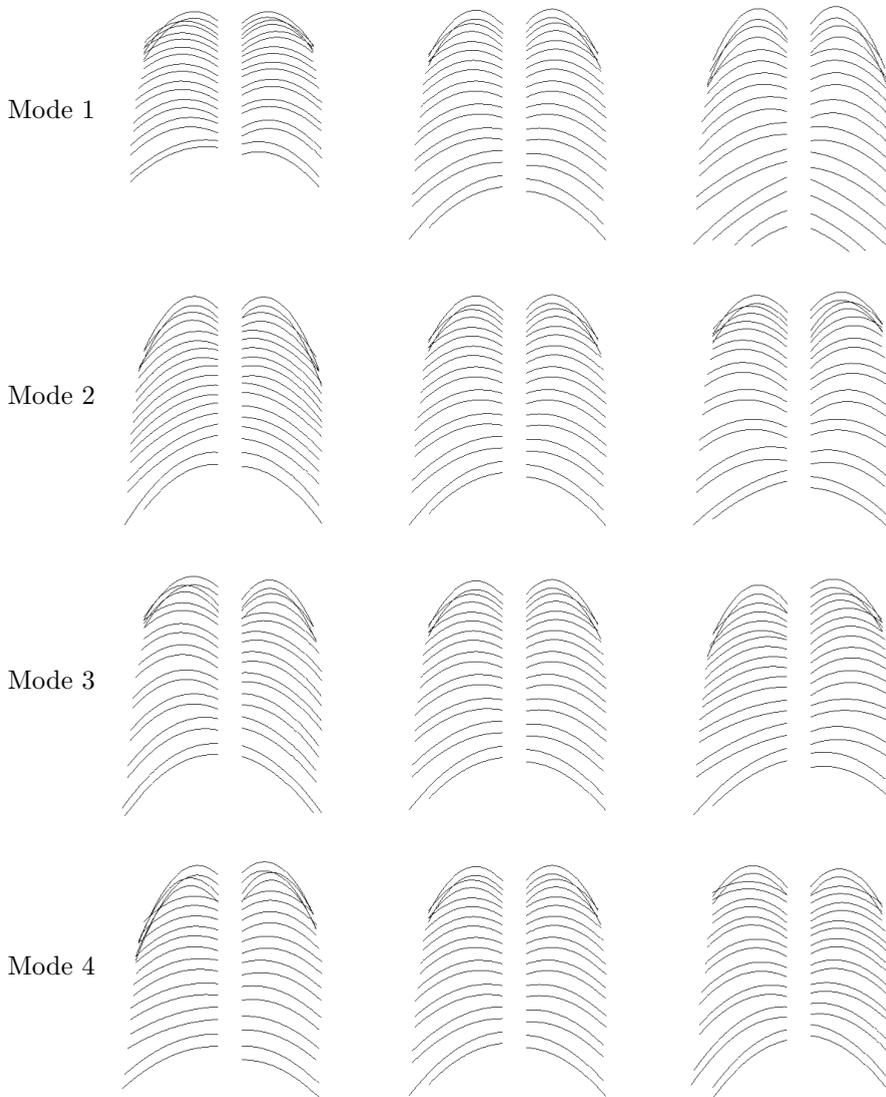


Figure 7.4: The average rib cage for all 35 images in the training set and the deformations of the rib cage for variations of the first 4 modes of the model from $-3\sqrt{\lambda_i}$ (left column) to 0 (middle column) to $+3\sqrt{\lambda_i}$ (right column).

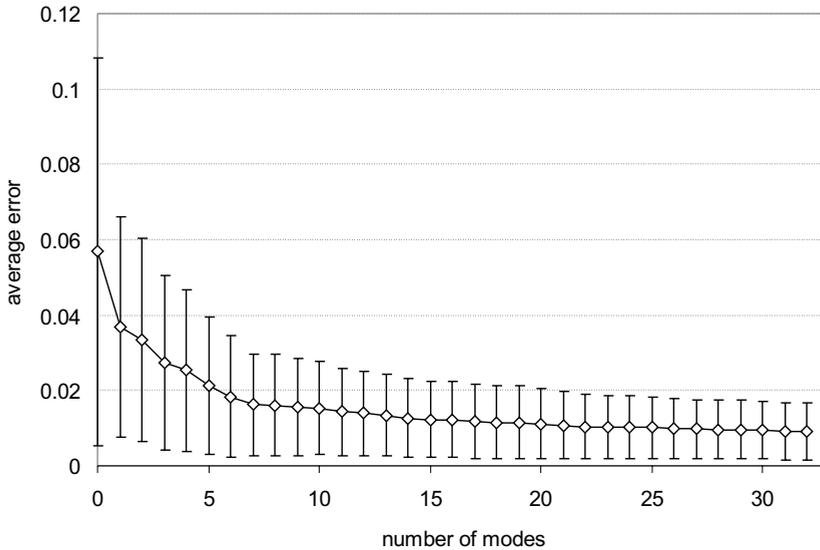


Figure 7.5: The graph shows the average distance and standard deviation over all images between the true landmark points and the fitted landmark points as a function of the number of modes used in the rib cage model. These results represent an upper boundary for the accuracy that can be reached by fitting the model to new data directly. The error distances are given in the lung coordinate system units (see Figure 7.2). An error of 1 corresponds roughly to 100 pixels (for images of 512×512 pixels).

model directly to the manual segmentation. This result indicates how well the shape model can approximate the manual segmentation and provides an upper bound for the result that can be obtained by a fitting procedure.

Rib cage fit measure

Rib borders usually coincide with edges in the image. The edge strength in a given direction α can be computed from the first order derivatives of the image in the x and y -direction for a particular scale σ . For a given set of model parameters \mathbf{b} a rib cage fit measure can be computed as follows. A rib cage is generated using the model parameters and pixels at the upper rib borders and lower rib borders are computed, at a certain fixed horizontal distance. We took a spacing of 4 pixels for the images of 512×512 pixels used in this study. For each rib border pixel, the direction α perpendicular to the rib border, pointing from the rib into the intercostal space, is computed using the two neighboring pixels. The edge strength in this direction is evaluated and all these strengths are added to give the total rib cage fit measure.

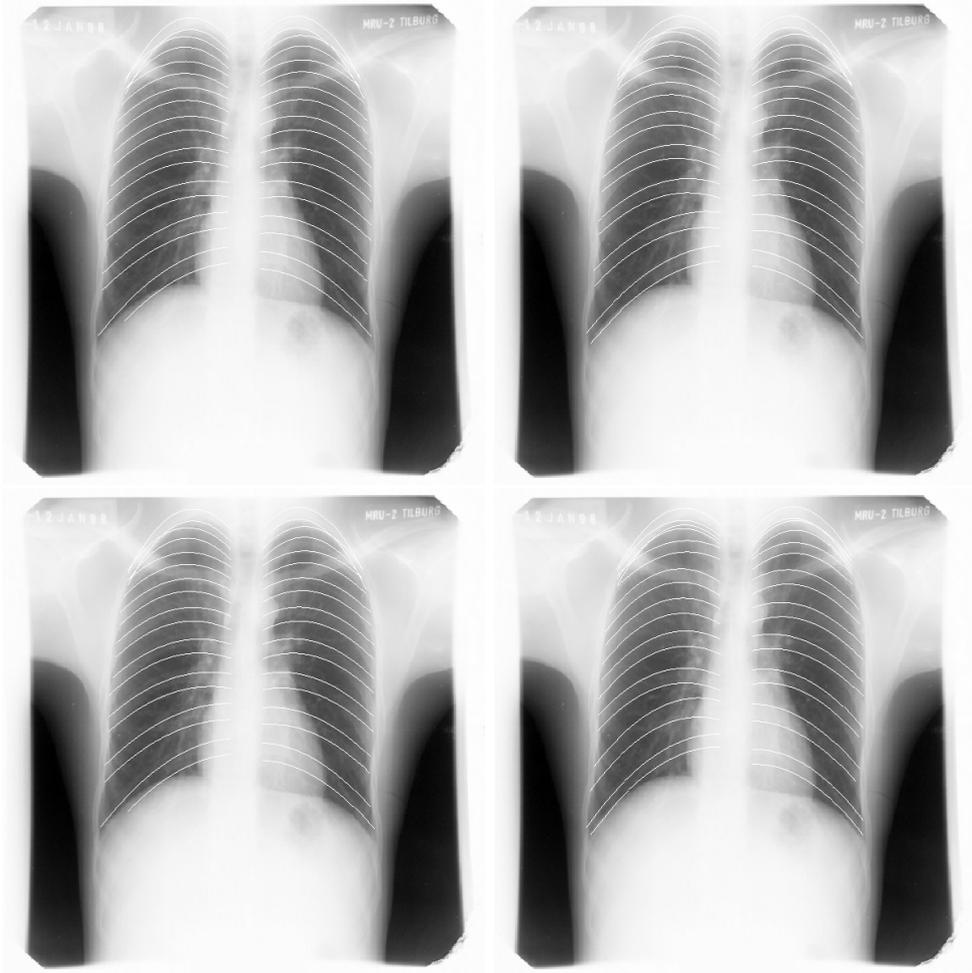


Figure 7.6: Results of fitting the rib cage model directly to the 72 points on the landmark lines. Top-left: the result of fitting with 0 modes in the model, which amounts to plotting the mean rib cage on the lung coordinate system of the image. Top-right: fitting the model with 3 modes $(-0.178, 0.493, -0.953)$. Bottom-left: fitting with 10 modes $(-0.178, 0.493, -0.953, -0.717, -1.08, -1.66, 0.580, -0.267, 0.615, -0.0902)$. Bottom-right: the true (manually segmented) rib cage.

Fitting the rib cage

The statistical rib cage is model is fitted directly to each image of the training set (with this training image left out of the model). We start with the mean rib cage (all model parameters set to 0). Powell's direction set method [202] is used to optimize the rib cage fit measure. The method requires about 4 seconds on a standard PC

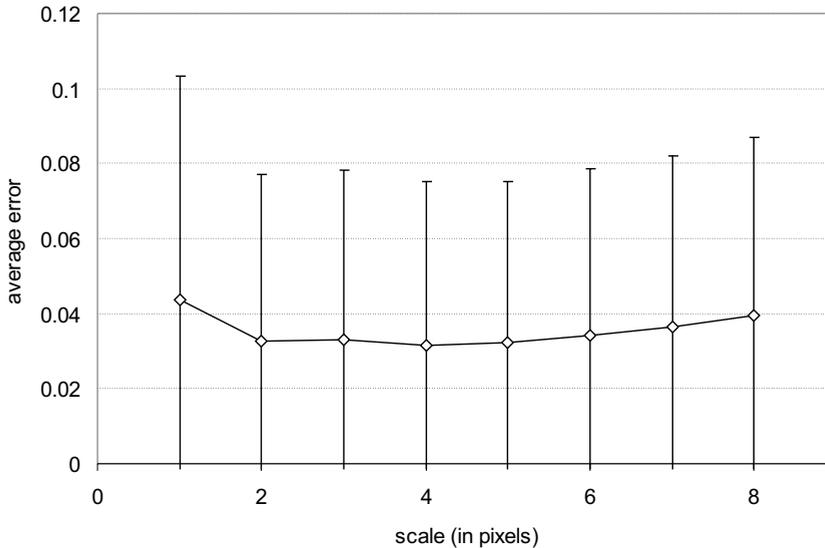


Figure 7.7: Total average error with standard deviation over all images, expressed in the scaled y -coordinate system introduced in Figure 7.2, for fitting the rib model with 10 modes to all images, as a function of scale σ .

with a 500 MHz processor. The only free parameters in this procedure are the scale σ with which the image derivatives are computed and the number of modes t in the model.

Figure 7.7 shows the error using a fixed number of 10 modes (which explains 98.3% of the total variance) as a function of the scale σ . The optimum scale appears to be $\sigma = 4.0$ pixels (images are 512×512 pixels). At this scale, the average error is 0.0316. This cannot be directly converted to pixels since the scaling is different for each image, but it amounts to about 3 pixels. It is evident from the flat slope of the graph in Figure 7.7 that the method is not very sensitive to the particular choice of σ .

Using this optimal scale, the model was fitted to all images for various numbers of modes. Figure 7.8 shows the error as a function of the number of modes, together with the error obtained by fitting the model directly to the true rib cage points, which is a theoretical upper bound for the accuracy that can be obtained.

In Figure 7.9 the segmented rib cages are shown using the optimal scale $\sigma = 4.0$ and 10 modes for the image with the lowest and highest error and the two images in between.

There are other ways to measure the performance of the method. One is to use the method to classify areas within the lung fields as costal or intercostal. The method with $\sigma = 4.0$ and 10 modes yields an accuracy of 0.787 ± 0.0873 . The best result obtained is 0.882, the worst is 0.587. In Figure 7.10 some of these results are shown.

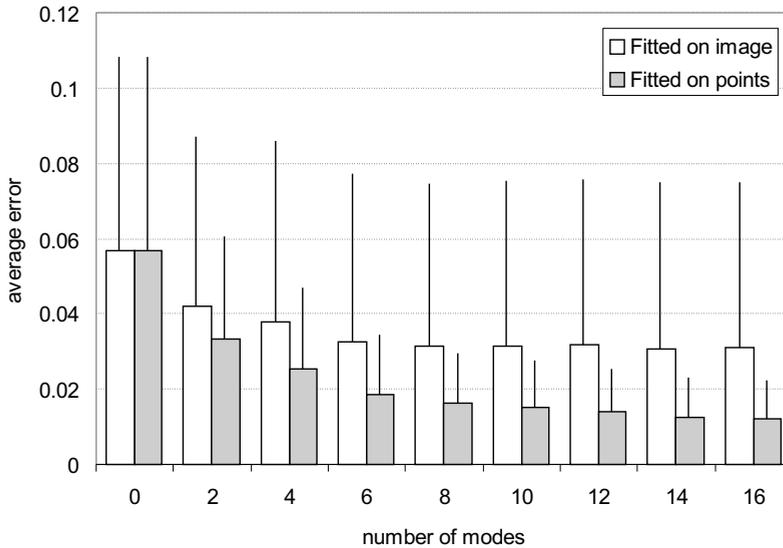


Figure 7.8: Total average error with standard deviation over all images, expressed in the scaled y -coordinate system introduced in Figure 7.2, for fitting the model to all images, using $\sigma = 4.0$ pixels and a variable number of modes. The results are shown in comparison to the theoretically optimal result of fitting the model directly to the rib cage points, as in Figure 7.5.

One of the key advantages of the statistical rib model presented here is that it captures correlations between all ribs. This makes it possible to infer the locations of parts of the rib cage from the locations of other parts in an optimal way. To demonstrate this capability, we show some examples of fitting the left part of the rib cage (located on the right of the image!) using only image information from the right lung field and vice versa. Representative results for two images are shown in Figure 7.12. For the image in the top row of Figure 7.12, the average errors for all three cases are shown per rib in Figure 7.11.

Discussion

We have shown how to construct a statistical model of the complete rib cage and how to fit this model to new images. The results are quite satisfactory, both in terms of robustness and accuracy. Only in a few cases, the algorithm failed to converge to a reasonable result. This might be improved by using a more sophisticated rib cage fit measure. Improving the objective function appears to be the most appropriate direction for future research because the current fit measure leads to unwanted behavior in some cases. For instance, it may force the upper border of the lowest ribs to coincide

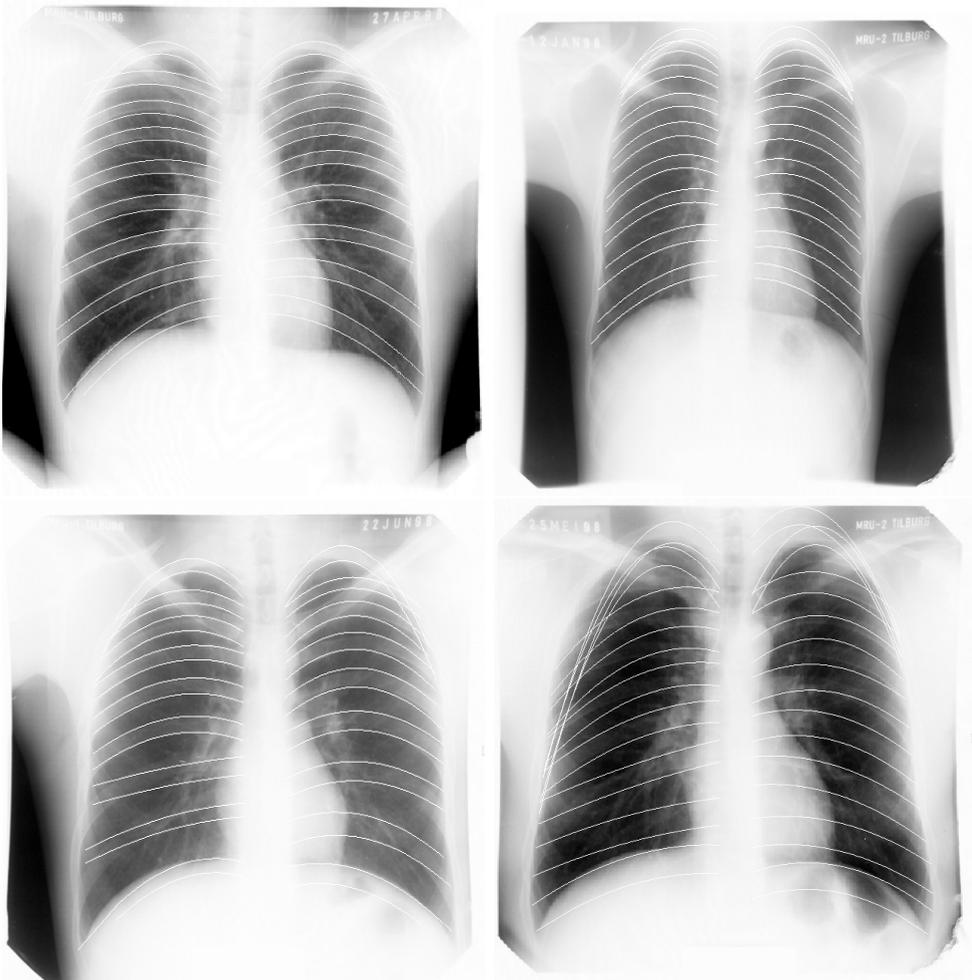


Figure 7.9: The rib cage model was fitted to all 35 images with $\sigma = 4.0$ pixels and 10 modes. The results were sorted on accuracy. Here results are shown for the images on top of the list with the smallest error (top-left), the image one-third down the list (top-right), two-third down the list (bottom-left) and the image at the bottom of the list with the largest error (bottom-right).

with the diaphragm (a strong edge), as can be seen from the result of the third image in Figure 7.9. Obviously the edges produced by rib border are not so strong for the lower rib as the edge produced by the diaphragm. Thus it would make sense to model the local structure caused by rib borders statistically.

Another open question is whether Powell's optimization method is appropriate to find a good or even the global maximum. Because of the periodicity in the rib cage, optimization methods may skip a rib. For example, if in the mean rib cage from which



Figure 7.10: Results of segmenting the lung fields into costal and intercostal area based on the delineated rib cage. The rib cage model was fitted with 10 modes and $\sigma = 4.0$ pixels. Black is correctly classified costal area (true positive). White is correctly classified intercostal area (true negative), or area outside lung fields. Dark gray is false negative, light gray is false positive. Left: image with best result (0.882 correct). Middle: image with median result (0.815 correct). Right: image with poorest result (0.587 correct). (Note that the images on the left and right are the same as the top-left and bottom-right images in Figure 7.9.)

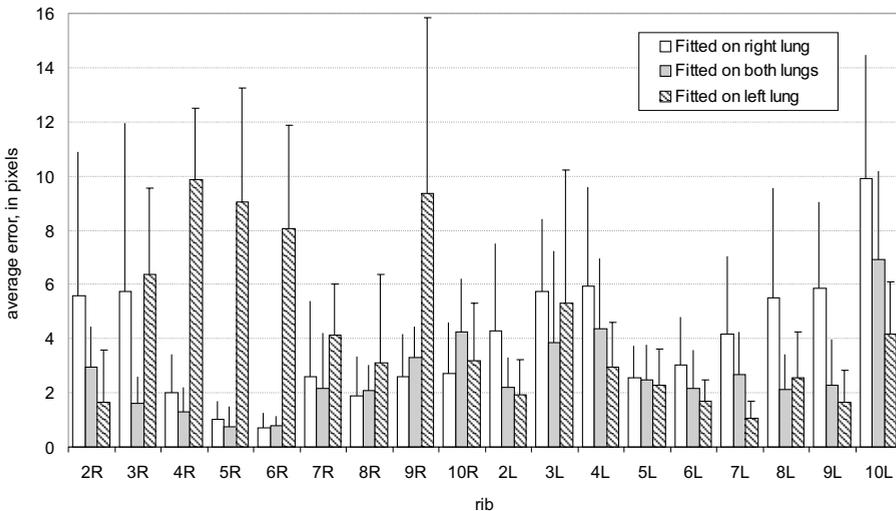


Figure 7.11: Average distance, with standard deviation, in pixels, between segmented rib border pixels and true rib border pixel for a single image (the top row in Figure 7.12), fitted with the rib cage model with 10 modes and $\sigma = 4.0$ pixels.

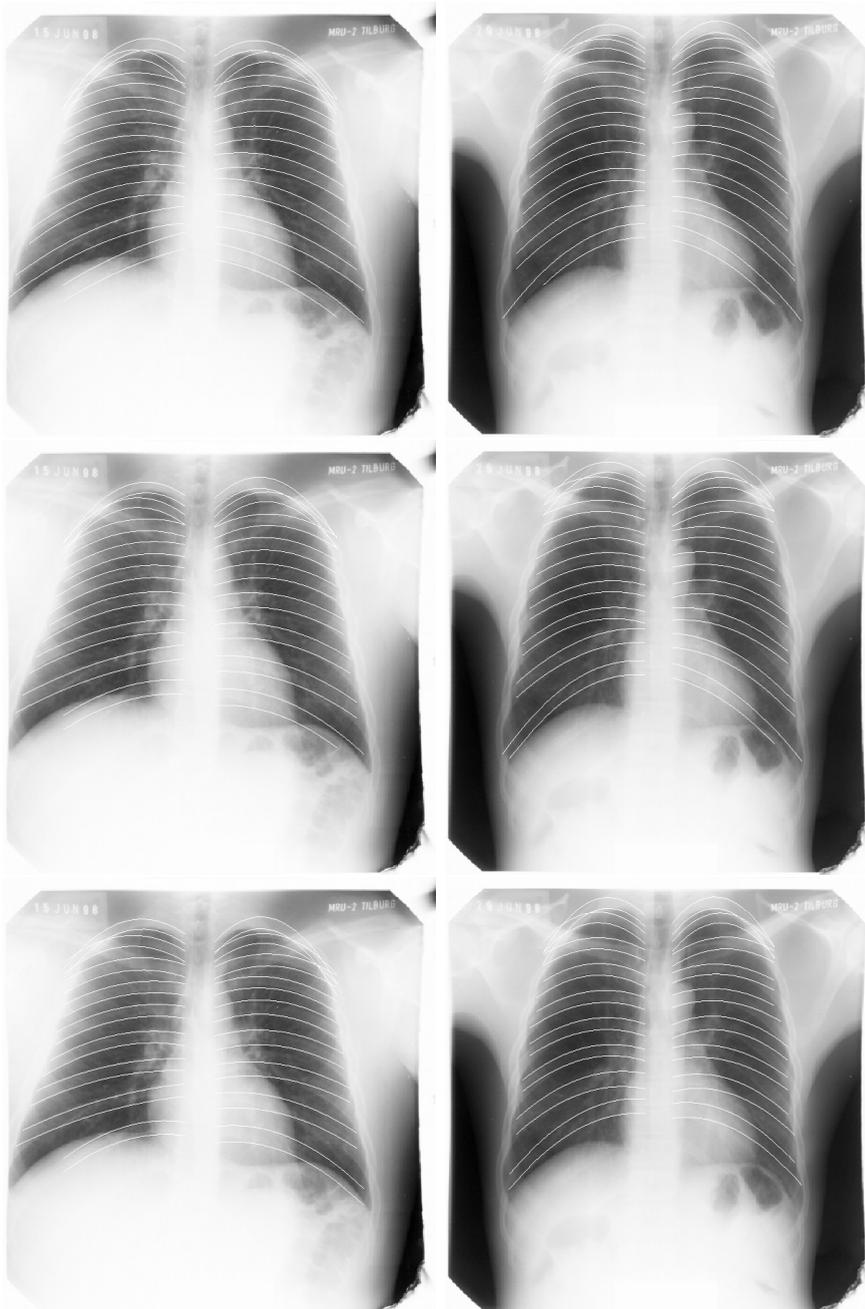


Figure 7.12: Results of fitting the rib cage model with 10 modes and $\sigma = 4.0$ pixels on two images, using only the ribs in the left lung to fit to (top row), using all ribs (middle row), and using only the ribs in the right lung (bottom row).

the optimization is started, the border of rib 4 is closer to the real location of rib 5, rib 4 may get stuck on rib 5 and another rib, in most cases rib 3, is skipped. Other optimization methods, such as simulated annealing, genetic algorithms, or taboo search may be less sensitive to this periodicity, which is likely to cause local maxima in the optimization space.

Even if the method gives good results, the ribs are often a few pixels away from their optimal location. This is an understandable weakness of the method as it stands. We use only about 10 parameters to deform the complete rib cage. Although these parameters capture almost all variations of the training data, each of the 18 ribs still has some degrees of freedom (say a slight deviation in translation, scaling, orientation) relative to all other ribs. Obviously, 10 model parameters cannot account for these effects. Using more modes would not solve this problem, as can be seen from Figure 7.8. If more training images were available, the explanatory power of a model with more modes could partly overcome this limitation. But it could also turn out that principal components cannot accurately describe rib cages.

A logical next step would be to use a fine-tuning procedure that allows each rib independently to move slightly to an optimal location. Note that the rib cage model can be used for this by considering only a single rib in the fitting procedure. In this refinement step, a more sophisticated rib model could be employed than that of two equidistant parabolas. The choice for this model, made because of its simplicity, forces the algorithm, and the person performing the manual segmentation as well, to make a compromise for each rib because usually not all parts of the upper and lower rib border can be accurately described by two equidistant parabolas. This effect increases the errors found in the fitting procedure. The active shape models segmentation method, possibly using optimal features as described in Chapter 6, seems well suited to find the optimal location of single rib using the initialization provided by the current method.

The major advantage of the current approach, compared to previous studies, is the fact that we present a single, global model for the complete rib cage, together with a straightforward fitting procedure. The result is a method that has very few free parameters and ad hoc choices. And these parameters and choices – the parabolic rib model, the lung reference coordinate system, the number of modes to retain in the model and the scale of the edge detection – are not critical; they merely tune the accuracy of the complete system versus the required computation time.

Chapter 8

Texture Analysis in Chest Radiographs

A fully automatic method is presented to detect abnormalities and to compute an abnormality score for a PA chest radiograph. The method is aimed at finding abnormal signs of a diffuse, textural nature, such as they are encountered in mass chest screening against tuberculosis.

The scheme starts with automatic segmentation of the lung fields, using Active Shape Models. The segmentation is used to subdivide the lung fields into overlapping regions of various sizes. Texture features are extracted from each region, using the moments of responses to a multi-scale filter bank. Additional difference features are obtained by subtracting feature vectors from corresponding regions in the left and right lung fields. A separate training set is constructed for each region. All regions are classified by voting among the k nearest neighbors, with leave-one-out. Next, the classification results of each region are combined, using a weighted multiplier in which regions with higher classification reliability weigh more heavily. This produces an abnormality score for each image.

The method is evaluated on two databases. The first database was collected from a TB mass chest screening program, from which 147 images with textural abnormalities and 241 normal images were selected. Although this database contains many subtle abnormalities, the classification has a sensitivity of 0.86 at a specificity of 0.50 and an area under the ROC curve of 0.820. The second database consists of 100 normal images and 100 abnormal images with interstitial disease. For this database the results were a sensitivity of 0.97 at a specificity of 0.90 and an area under the ROC curve of 0.986.

DESPITE THE EXISTENCE of a cheap cure, tuberculosis (TB) is a leading killer of adults in the world [193]. Mass chest screening can identify cases of active TB to fight the epidemic. The advent of digital chest units could facilitate the application of computer-aided diagnosis to improve the efficiency of mass chest screening. TB may reveal itself in many different radiographic patterns, but in most cases a chest radiograph of a patient with TB contains areas with diffuse

abnormalities. In this work we propose a scheme to detect such textural abnormalities.

The important role of texture analysis in computer-aided diagnosis is increasingly recognized [234]. Computer-aided diagnosis (CAD) for chest screening against TB is a new area that has not been directly investigated before. However, the design of CAD schemes for interstitial disease in general has received considerable attention (see also Chapter 2). Early work dates from the 1970s and focuses on the detection of pneumoconiosis, using features extracted from pixel profiles, Fourier spectra and co-occurrence matrices [227, 205, 142, 97, 115, 116, 237, 114]. More recent methods are applied to interstitial disease in general, and essentially use a similar approach: regions of interest (ROIs) are (automatically [36]) selected and texture features are computed, based on the Fourier spectrum [126, 127, 185, 184], geometric features [129], responses to filters [133], pixel profiles [111], fractal dimension [134], or combinations of feature sets [128]. Classification is performed using rules (thresholds) or feed-forward neural networks.

Texture analysis is an active research field and a large number of schemes for texture feature extraction have been proposed [100, 87, 204]. We have developed a texture feature extraction method based on a scale-space description of local histograms (see Chapter 5). This method combines and generalizes filter bank methods, *e.g.* [172, 24, 150, 241, 243] and co-occurrence matrices [101]. Experiments indicated that the filter bank method, in which the image is filtered and moments of local histograms are used as texture features, outperformed methods based on co-occurrence matrices and features extracted from the Fourier spectrum. This motivates our choice to use features extracted from the histograms of responses of regions in the image to a multi-scale filter bank consisting of the Gaussian filter and its derivatives.

A main problem in the texture analysis of chest radiographs is the complex “background” of superimposed normal anatomical structures to which the analysis must be somehow insensitive. One way to solve this problem would be to restrict texture analysis to regions of interest that do not contain normal structures such as (crossing) rib borders and (large) vessel projections [36, 127]. An alternative could be to pre-process the images so as to remove normal background structures [155, 253], but such approaches have not yet been applied to texture analysis of lung fields, although localized background structure in relatively small ROIs have been removed by use of a background trend correction technique [126]. Our approach is to divide the lung fields in parts and analyze each part separately, with a classifier trained with texture features extracted solely from these parts. In this way, the classifier should capture knowledge regarding the normal variation within that particular part. Overlapping regions of various sizes are used so that not only the texture features are extracted at multiple scales, but also from multiple apertures. Thus we perform *local* texture analysis with multiple *inner* scales and *outer* scales.

Materials

TB database

The TB database contains 326 abnormal and 290 normal posterior-anterior (PA) chest radiographs collected from a tuberculosis screening program for people seeking political asylum in The Netherlands. The abnormal images are representative for abnormal images encountered in mass chest screening: they consist of all abnormalities – as far as they could be traced – from a consecutive period of 18 months of screening, during which approximately 25,000 examinations were made.

All images were read by two radiologists. If one of them – or both – considers the image to be abnormal, the patient is contacted for further examination. The chest units used in this screening program are mobile Oldelft Electrodelca units (Nucletron BV, Veenendaal, The Netherlands). The tube voltage was 117 kV and the images were printed on 10 by 10 centimeter film. These films were digitized with a Lumisys 100 scanner (Lumisys, Inc., Sunnyvale, CA) to 932 by 932 pixels (pixel size 0.1 mm; a small area at the sides of the film was cropped since it was not exposed) with 12 bit intensity.

In order to be able to train classifiers to distinguish between normal and abnormal areas, two radiologists have outlined the area containing abnormalities in each abnormal image (each radiologist did this for half of the total set of abnormal images). During this procedure, several abnormal images were re-classified by the radiologist as normal. These images have been excluded from the set, leaving a total of 279 abnormal images.

The ground truth that we are trying to reproduce by computer for this database is the judgement of the radiologists who read the films. In a large number of cases (106 out of the 279 abnormal cases), one of the radiologists judged the image to be normal, while the other favoured to contact the patient for further examination. We consider these images as abnormal.

TB is known to often affect the upper lung fields. Figure 8.1(a) gives an idea of the spatial distribution of the abnormal areas within the chest radiographs in the TB database.

Not all abnormalities are suitable for detection by texture analysis. Each abnormal image was assigned to one or more of the following categories:

1. Shape abnormalities. In these images the shape of the lung regions or other objects in the images was abnormal. This category contains images with enlarged heart size, unusual heart shape, a blunt sinus, broadened mediastinum, and unusual lung shapes because parts of the lungs have been destroyed or are filled with fluid. This category contains 36 images.
2. Densities/objects. In many cases the abnormal area is homogeneously denser compared to its surroundings without showing a clear textural difference. The diameter of the densities ranges in size from a few millimeters to several centimeters. This category contains 66 images.
3. Textural abnormalities. The size of abnormal areas ranges from a few square centimeters in diameter to the complete lung fields. This was the largest class of

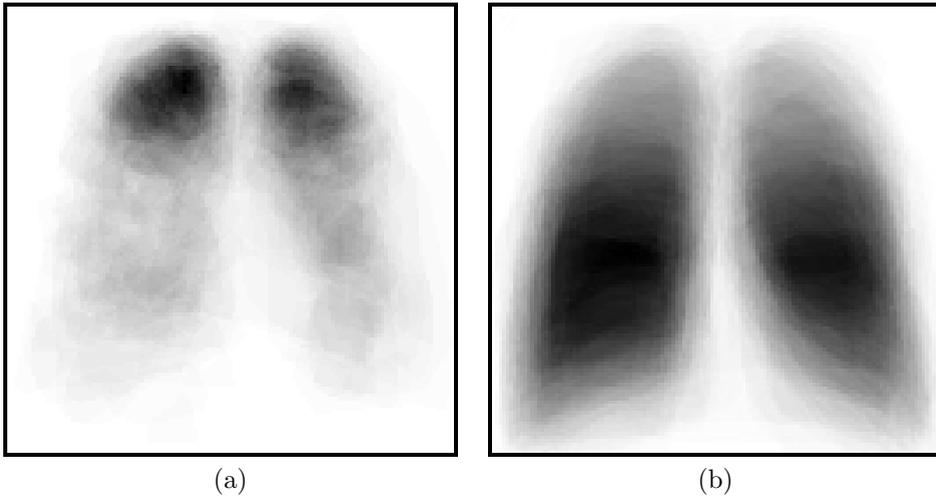


Figure 8.1: (a) Distribution of abnormal areas for the TB database. This image is computed by superimposing binary images (with abnormal areas in black) containing the abnormal areas outlined by the radiologist of all 279 abnormal images. It can be seen that the upper lung fields are more often abnormal than the middle and lower lung fields, with the right lung top being slightly more often abnormal than the left lung top. (b) Distribution of abnormal areas for the ID database. The image is produced in the same way, using abnormal areas from the 100 abnormal cases. Compared to the TB database, the abnormal areas are spread more evenly over the lung fields with a preference for the lower lung fields.

abnormal images. There is considerable variation in the character and subtlety of the abnormal textures. This category contains 166 images.

4. Hilar regions. Radiologist are particularly suspicious of enlarged hila since these may indicate the presence of TB. This category contains 13 images.
5. Small abnormalities in the lung top. Usually these are small fibrotic lesions and often they are not harmful but in some cases patients are contacted for further examination. The abnormal area is very small in these cases. This category contains 24 images.
6. Small pleural abnormalities at the diaphragm. This category contains 15 images.

People are photographed fully dressed in this screening. This results in a large number of images with artifacts from clothing: buttons, zippers, clips, and a variety of other objects.

For this study, all normal cases and only those abnormal cases with textural abnormalities were used. Furthermore, all images with large clothing artifacts were removed. (This includes necklaces and buttons in the lung fields. Images containing artifacts outside the lung fields and small artifacts such as brassiere clips were not

removed.) The resulting database contained 147 abnormal cases and 241 normal cases.

ID database

This database contains 100 normal PA chest radiographs and 100 abnormal PA chest radiographs with interstitial disease and is referred to as the ID database. The images were obtained at the University of Chicago Hospitals. All normal cases were selected based on consensus of independent review on each radiograph by four experienced chest radiologists. The abnormal cases, which ranged from subtle to severe, were selected based on radiological findings, CT, clinical data and/or follow-up radiographs, by consensus of the same radiologists. Some images contain artifacts due to clothing or catheters. The radiographs were digitized to 2000 by 2000 pixels with 0.175 mm pixel size and 10 bits intensity.

A chest radiologist classified each upper, middle and lower lung field of the abnormal cases as normal, possibly abnormal and definitively abnormal. For this study, the possibly and definitively abnormal parts were used to obtain abnormal areas for training classifiers. Figure 8.1(b) gives an idea of the spatial distribution of the abnormal areas within the chest radiographs.

Methods

Overview

The system consists of 4 components:

- Segmentation
- Texture analysis
- Classification of texture feature vectors
- Classification of the image

Segmentation

Active Shape Model segmentation

The purpose of the segmentation is to find corresponding regions within the lung fields. Segmentation of lung fields on PA chest radiographs has received considerable attention in the literature (see also Chapter 2). Rule-based schemes have been proposed by Armato *et al.* [6], Xu *et al.* [269, 270], Duryea and Boone [65], Pietka [198], and Brown *et al.* [28]. Lung segmentation by pixel classification using neural networks has been investigated by McNitt-Gray *et al.* [178], and Tsujii *et al.* [235]. Vittitoe *et al.* [250, 251] developed a pixel classifier for the identification of lung regions using Markov random field modeling. Van Ginneken and ter Haar Romeny proposed a hybrid method that combines a rule-based scheme with a pixel classifier [84, Chapter 3].

In this work the lung fields are delineated with the Active Shape Models (ASMs) method introduced by Cootes and Taylor. This method is presented and discussed in Chapter 6. To make this chapter self-contained, we briefly state the relevant formulas and symbols from the ASMs method here. For details, see Chapter 6 and [49] or [48].

An object is described by n points, referred to as landmark points. The landmark points are manually determined in a set of s training images. From these collections of landmark points, a point distribution model is constructed. The landmark points $(x_1, y_1), \dots, (x_n, y_n)$ are stacked in shape vectors $\mathbf{x} = (x_1, y_1, \dots, x_n, y_n)^T$ and principal component analysis is applied to a collections of s such shapes (without performing any kind of alignment), by computing the mean shape and the covariance,

$$\bar{\mathbf{x}} = \frac{1}{s} \sum_{i=1}^s \mathbf{x}_i, \quad (8.1)$$

$$\mathbf{S} = \frac{1}{s-1} \sum_{i=1}^s (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \quad (8.2)$$

and the eigensystem of the covariance matrix. The eigenvectors corresponding to the t largest eigenvalues λ_i are retained in a matrix $\Phi = (\phi_1 | \phi_2 | \dots | \phi_t)$. A shape can now be approximated by

$$\mathbf{x} \approx \bar{\mathbf{x}} + \Phi \mathbf{b} \quad (8.3)$$

where \mathbf{b} is a vector of t elements containing the model parameters, computed by

$$\mathbf{b} = \Phi^T (\mathbf{x} - \bar{\mathbf{x}}). \quad (8.4)$$

When fitting the model to a set of points, the values of \mathbf{b} are constrained to lie in the range of m times $\pm\sqrt{\lambda_i}$.

The number t of eigenvalues to retain is chosen so as to explain a certain proportion f_v of the variance in the training shapes. The desired number of modes is given by the smallest t for which

$$\sum_{i=1}^t \lambda_i \geq f_v \sum_{i=1}^{2n} \lambda_i. \quad (8.5)$$

To fit the shape model to an image, profiles $\mathbf{g}_1 \dots \mathbf{g}_s$ are sampled (using linear interpolation) around each landmark, perpendicular to the contour. Sampling k pixels on either side of the profiles gives profiles of length $2k + 1$. The first derivatives of these profiles are used. The profiles are also normalized, by dividing through the sum of absolute values of the elements. For each landmark point, the mean profile $\bar{\mathbf{g}}$

and the covariance matrix \mathbf{S}_g are computed. This allows for the computation of the Mahalanobis distance between a new profile \mathbf{g}_i and the profile model

$$f(\mathbf{g}_i) = (\mathbf{g}_i - \bar{\mathbf{g}}) \mathbf{S}_g^{-1} (\mathbf{g}_i - \bar{\mathbf{g}}) \quad (8.6)$$

These profile models, given by $\bar{\mathbf{g}}$ and \mathbf{S}_g , are constructed for multiple resolutions. The finest resolution uses the original image and a step size of 1 pixel when sampling the profiles. The next resolution is the image observed at scale $\sigma = 1$ and step size of 2 pixels. Subsequent levels are constructed by doubling the image scale and the step size.

Shapes are fitted in an iterative manner, starting from the mean shape. Each landmark is moved along the direction perpendicular to the contour to n_s positions on either side, evaluating a total of $2n_s + 1$ positions. The landmark is put at the position with the lowest Mahalanobis distance. After moving all landmarks, the shape model is fitted to the points, yielding an updated segmentation. This is repeated a fixed number of N times at each resolution, from coarse to fine.

Subdividing the lung fields

To analyze smaller areas, the lung fields are subdivided into smaller regions. As a first step, each lung field is divided into an upper, middle and lower part. This is done for a hypothetical chest image with the lung fields of the training images at their mean location, by computing horizontal lines that divide the lung fields in three parts of equal area (Figure 8.2, right image). These parts are subdivided into four areas of equal area, by first subdividing them vertically (Figure 8.2, middle image) and then horizontally (Figure 8.2, left image). In this way, a total of 42 regions are defined.

After segmentation, the region maps as shown in Figure 8.2 are warped to the segmentation result, using interpolation with radial basis functions [209]. After warping, the borders between the regions need no longer be horizontal or vertical.

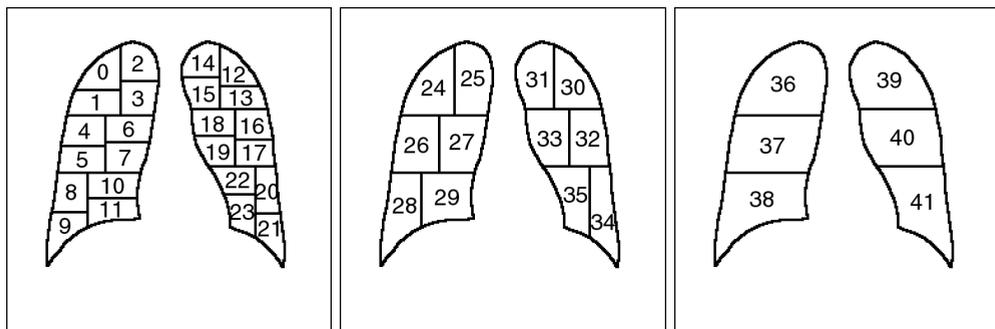


Figure 8.2: The lung fields are subdivided into 24 regions. These regions are merged to form larger regions. In total 42 regions, numbered 0 to 41, are used. This numbering scheme is used throughout this chapter. The shape and location of the lung fields shown here are obtained from the mean lung fields of the manually segmented training images.

Finally, a criterion is needed that determines when a region is abnormal. The following rule is adopted: if there is any overlap between a region and an abnormal area drawn by the radiologist, the region is considered to be abnormal.

Texture analysis

We use a multi-scale filter bank to extract texture features. Filters are Gaussian derivatives $L_n^\theta(x, y; \sigma)$ where σ denotes the scale, n the order of derivative and θ the direction in which the derivative is computed.

The images in the TB database were subsampled to a working resolution of 700 by 700 pixels and 6 filters were used, L_0 , $L_1^{0^\circ}$, $L_1^{90^\circ}$, $L_2^{0^\circ}$, $L_2^{60^\circ}$, $L_2^{120^\circ}$, at 4 scales $\sigma = 1, 2, 4, 8$ pixels. The images in the ID database were subsampled to 1400 by 1400 and the same filters were used with scales $\sigma = 1, 2, 4, 8, 16$ pixels.

The second, third, and fourth moment (standard deviation, skew and kurtosis, respectively) of each filtered image, averaged over the region under consideration, are computed as texture features.

These moments are independent of the mean density in each region. Because the density may also be an indicator of abnormality (abnormal regions being more dense), the mean density of each region is also computed as a texture feature. Furthermore, the mean density of each region after normalizing the density of the whole image, and after normalizing the density in the lung fields, is also used.

When reading a chest radiograph, differences between corresponding regions on the left and right lung fields are an important clue for radiologists, since most abnormalities are asymmetrical. This idea can be implemented in a straightforward way for the generation of extra features by subtracting feature vectors between corresponding regions in the left and right lung field. These features are referred to as “difference features”. To be able to do this, the left lung field was flipped before computing texture features.

Classifying texture feature vectors

Texture analysis yields a texture feature vector for each region in each image. The next step is to estimate the abnormality of each region based on these feature vectors. To this end, a training dataset is constructed for each region. This set contains the feature vectors of all abnormal regions, to which an equal number of normal feature vectors, from randomly selected normal images, are added. For each feature, scaling factors are computed to normalize the mean to zero and the variance to one. The same scaling factors are applied to feature vectors before classification.

The k nearest neighbors of the feature vector are extracted from the training set, leaving out the feature vector to be classified, if necessary (which is easy to implement by simply ignoring neighbors at zero distance whenever they occur). Each neighbor votes for the region to be normal or abnormal. We used a fast algorithm for finding the k nearest neighbors developed by Arya and Mount [8], available on the web at <http://www.cs.umd.edu/~mount/ANN>. Instead of a binary normal/abnormal decision when classifying feature vectors, a probability measure that a region is abnormal is computed using weighted voting among the k nearest neighbors. If there

are k neighbors at distances d_1, \dots, d_k from the feature vector to be classified, and c_1, \dots, c_k are the classes of the k neighbors, with $c = 0$ for normal and $c = 1$ for abnormal regions, the classification for the input sample is

$$c = \frac{\sum_{i=1}^k \exp\left(\frac{-d_i^2}{\sigma^2}\right) c_i}{\sum_{i=1}^k \exp\left(\frac{-d_i^2}{\sigma^2}\right)}, \quad (8.7)$$

where σ is the mean of d_1, \dots, d_k .

The classification is a number in the range from 0 (normal) to 1 (abnormal). Given these classifications, ROC analysis can be performed. An ROC curve plots the true positive fraction as a function of the false positive fraction. The area under the ROC curve, indicated as A_z , indicates how reliable the classification can be performed for the region under consideration [181]. A value of $A_z = 1$ represent perfect classification, $A_z = 0.5$ corresponds to random guessing. A_z and its standard deviation are computed by fitting a bi-normal model to the ROC data. Software to do this is available at <http://www-radiology.uchicago.edu/sections/roc>.

Classifying images

The classification results per region may be a useful final result for a CAD scheme. Regions with a high abnormality score could be indicated while a radiologist is reading the image. In this work we aggregate the regional scores into a stand-alone result: an estimate of the complete image being normal or abnormal. In principle, the results of each region could be used directly as features to classify the image as a whole – using any classifier. Here we propose a classifier that *multiplies* the probabilities that regions are normal. The motivation for this approach is that the image is abnormal if *any* of the regions is abnormal. By multiplying the probabilities that the regions are normal we obtain zero probability that the image is normal if any of the regions has zero probability of being normal. The area A_z under the ROC curve for each region – an indicator of the classification performance of the scheme for that particular region – is used to weigh the probabilities for each region. If A_z is below some threshold T_{A_z} the region is not taken into account, if $A_z = 1$, the region is taken fully into account.

The weighted probability of an image being abnormal is given by

$$c = 1 - \prod_{i=1}^N (1 - f_i c_i), \quad (8.8)$$

where c_i denotes the classification of the i -th region which is in the range from 0 (normal) to 1 (abnormal), N is the number of regions, and f_i is the weighting factor for region i given by

$$f_i = \text{Max}\left(\frac{A_{zi} - T_{A_z}}{1 - T_{A_z}}, 0\right). \quad (8.9)$$

The threshold T_{A_z} determines the minimum value that A_{zi} , the area under the ROC curve for the i -th region, must have in order to be taken into account in the classification of the complete image.

Results

Segmentation

To train the Active Shape Model, the author manually outlined the lung contours in 230 chest images of the TB database, under supervision of a radiologist. These images were subsampled to a resolution of 256 by 256 pixels. From the outlines, 116 landmark points were extracted, 60 for the right lung and 56 for the left lung, in the following way. From each lung contour the highest point (lung apex), the sinus and the crossing of the diaphragm and the mediastinum are determined. These 6 fixed points become landmark points. The other landmarks points are obtained by equidistant sampling of points along the lung contour between these fixed points.

The threshold f_v was set at 0.99. This is a rather high value, but note that information about the size, position and orientation of the lung shapes is all contained in the data since no alignment was performed. This resulted in a shape model with 15 modes. When the shape model was fitted, the model parameters were limited to a range of $\pm 2\sqrt{\lambda_i}$.

To segment the lung fields, profiles with length 9 ($k=4$) and 5 levels of resolution were used. During fitting, the number of positions evaluated at each iteration for each landmark point was 9 ($n_s = 4$) and the number of iterations at each resolution $N = 10$.

The 230 training images are a subset from the TB database. When an image was segmented that was also part of the training set, a new model, both for the shape and the profiles was computed with all training images, except the one to be segmented (the leave-one-out approach).

In Figures 8.4, 8.5, and 8.6 at the end of this chapter the warped regions after segmentation for several normal and abnormal images from the TB database are shown. Table 8.1 lists the total number of abnormal regions in the TB and ID databases for each region.

Note that the shape models and profile models were constructed from 8images of the TB database and are used to segment the images in the ID database. This can lead to poor results if there are structural differences between these databases, such as differences in the shape and size of the lung fields, and/or differences in the gray level profiles perpendicular to the landmarks. With regard to the size of the lung fields there is certainly a structural difference: the lung fields are often larger in the ID database. However, for most images in the ID database, segmentation results were satisfactory. To qualitatively evaluate the segmentation results, the number of failures to accurately segment the contours along the outer rib cage, the diaphragm, the mediastinum and the lung top were counted by the author. The results are given in Table 8.2. A failure is defined as a segmentation in which at least one of the 24 smallest regions into which the lung fields are divided (Figure 8.2) is 50% incorrect.

Texture analysis and classification

For the classification of texture vectors computed from regions, a value of k nearest neighbors must be chosen. We empirically selected $k = 7$. When combining the scores for each region into an aggregate score for the complete image, a threshold T_{A_z} must be chosen. We selected $T_{A_z} = 0.60$. Varying k between 3 and 9 and T_{A_z} between 0.5 and 0.75 had hardly any impact on the total classification results or the results per region: deviations were almost always within the standard deviation of A_z .

Selection of different subsets of features had far larger effects. Various subsets

TB database							
region	#	region	#	region	#	region	#
0	86	12	73	24	88	36	96
1	52	13	55	25	93	37	48
2	90	14	69	26	37	38	33
3	68	15	58	27	36	39	79
4	29	16	36	28	22	40	49
5	24	17	22	29	26	41	32
6	31	18	34	30	77		
7	19	19	25	31	74		
8	19	20	24	32	43		
9	19	21	18	33	39		
10	20	22	24	34	28		
11	23	23	25	35	30		

ID database							
region	#	region	#	region	#	region	#
0	50	12	43	24	72	36	72
1	72	13	60	25	70	37	96
2	46	14	35	26	95	38	95
3	70	15	57	27	94	39	61
4	86	16	74	28	94	40	84
5	94	17	84	29	95	41	89
6	83	18	70	30	60		
7	93	19	82	31	57		
8	94	20	89	32	84		
9	89	21	81	33	82		
10	95	22	89	34	89		
11	90	23	85	35	89		

Table 8.1: Top: the number of abnormal regions in the TB database, for each region. The total number of abnormal images is 147. Bottom: the number of abnormal regions in the ID database, for each region. The total number of abnormal images is 100. See Figure 8.2 for the region numbering scheme.

	TB database		ID database		All images
	normal	abnormal	normal	abnormal	
# of images	241	147	100	100	589
Right rib cage	3 (1%)	5 (3%)	0 (0%)	1 (1%)	9 (2%)
Left rib cage	4 (2%)	4 (3%)	0 (0%)	1 (1%)	9 (2%)
Right lung top	16 (7%)	18 (12%)	9 (9%)	24 (24%)	67 (11%)
Left lung top	19 (8%)	22 (15%)	8 (8%)	25 (25%)	74 (13%)
Right mediastinum	0 (0%)	1 (1%)	6 (6%)	1 (6%)	8 (1%)
Left mediastinum	5 (2%)	5 (3%)	6 (6%)	7 (7%)	23 (4%)
Right diaphragm	10 (4%)	9 (6%)	11 (11%)	13 (13%)	43 (7%)
Left diaphragm	19 (8%)	22 (15%)	14 (14%)	16 (16%)	71 (12%)

Table 8.2: A qualitative evaluation of the performance of the ASM segmentation of the lung fields. The number of failures to accurately segment those parts of lung contour corresponding to the rib cage, the lung tops, the mediastinum and the diaphragm, are listed here. A failure is defined as a result in which the segmentation of at least one of the 24 smallest regions into which the lung fields are divided is at least 50% incorrect.

were evaluated on both databases. In all cases we compared the original features, the difference features and the combination of original and difference features. The results are listed in Table 8.3.

For features sets containing only second moments and a single scale, $\sigma = 2$ and $\sigma = 4$ gave the best results on the TB database. The set of difference features outperformed the set of original features, as was the case for most feature sets. Performance slightly increased by using multiple scales. The highest A_z value of 0.820 ± 0.022 is obtained using the difference features from $\sigma = 2, 4$ and $m = 2$. Including higher moments did not improve the results. The performance of density features was rather poor.

The results for the ID database are quite different. First of all, the A_z values are much higher. They approach 1, the perfect result. Because the resolution of the ID images is higher, we used one more scale in the multi-resolution analysis. When using only second moment features, all scales lead to good results with only a decrease for $\sigma = 16$ compared to smaller scales. Apparently the structures that make up the difference between normal and abnormal lung texture are almost blurred away at this scale, just as is the case for the corresponding scale of $\sigma = 8$ in case of the TB images.

Interestingly and contrary to the results from the TB database, higher moments (skew and kurtosis) yield good results for the ID database. The difference features all have lower A_z values, again contrary to the experiments on the TB database. The best result, $A_z = 0.986 \pm 0.0062$, is obtained using the original features for three scales $\sigma = 1, 2, 4$, and $m = 2, 3, 4$. As in the experiments on the TB database, using density features does not improve performance.

For both databases, several features sets provide comparable results within the standard deviation. For the TB database, classification performance is around $A_z = 0.80$, for the ID database, classification performance is around $A_z = 0.98$.

Figure 8.3 plots the ROC curves for the best performing features sets for the TB

TB database					
feature set		original	difference	both	
σ	m	density			
1	2		0.743 ± 0.026	0.765 ± 0.024	0.730 ± 0.026
2	2		0.792 ± 0.024	0.812 ± 0.022	0.796 ± 0.023
4	2		0.766 ± 0.025	0.798 ± 0.024	0.813 ± 0.023
8	2		0.686 ± 0.027	0.777 ± 0.025	0.739 ± 0.026
2,4	2		0.789 ± 0.024	0.820 ± 0.022	0.812 ± 0.022
1,2,4	2		0.783 ± 0.024	0.800 ± 0.024	0.789 ± 0.023
1,2,4,8	2		0.767 ± 0.025	0.793 ± 0.024	0.787 ± 0.024
2,4	3,4		0.741 ± 0.026	0.672 ± 0.028	0.723 ± 0.026
2,4	2,3,4		0.773 ± 0.024	0.720 ± 0.027	0.770 ± 0.024
		×	0.589 ± 0.030	0.633 ± 0.029	0.629 ± 0.030
2,4	2	×	0.796 ± 0.023	0.789 ± 0.024	0.784 ± 0.024

ID database					
feature set		original	difference	both	
σ	m	density			
1	2		0.948 ± 0.015	0.820 ± 0.029	0.947 ± 0.015
2	2		0.948 ± 0.015	0.816 ± 0.030	0.944 ± 0.015
4	2		0.935 ± 0.018	0.795 ± 0.031	0.917 ± 0.020
8	2		0.941 ± 0.016	0.820 ± 0.029	0.929 ± 0.018
16	2		0.905 ± 0.021	0.825 ± 0.029	0.907 ± 0.021
1	3,4		0.960 ± 0.012	0.851 ± 0.027	0.941 ± 0.016
2	3,4		0.958 ± 0.013	0.846 ± 0.027	0.951 ± 0.014
4	3,4		0.956 ± 0.013	0.786 ± 0.032	0.938 ± 0.016
8	3,4		0.932 ± 0.017	0.758 ± 0.033	0.895 ± 0.022
16	3,4		0.912 ± 0.019	0.827 ± 0.029	0.909 ± 0.020
1,2	2		0.950 ± 0.014	0.824 ± 0.029	0.948 ± 0.014
1,2,4	2		0.954 ± 0.013	0.832 ± 0.028	0.949 ± 0.014
1,2,4,8	2		0.956 ± 0.013	0.849 ± 0.027	0.953 ± 0.014
1,2,4,8,16	2		0.957 ± 0.013	0.860 ± 0.026	0.949 ± 0.014
1,2	2,3,4		0.983 ± 0.0073	0.872 ± 0.024	0.980 ± 0.0079
1,2,4	2,3,4		0.986 ± 0.0062	0.866 ± 0.025	0.977 ± 0.0087
1,2,4,8	2,3,4		0.986 ± 0.0062	0.846 ± 0.027	0.973 ± 0.0097
1,2,4,8,16	2,3,4		0.984 ± 0.0071	0.862 ± 0.026	0.968 ± 0.011
		×	0.745 ± 0.034	0.695 ± 0.036	0.787 ± 0.032
1,2,4,8	2,3,4	×	0.983 ± 0.0071	0.847 ± 0.027	0.970 ± 0.011
1,2,4,8,16	2,3,4	×	0.983 ± 0.0072	0.856 ± 0.026	0.968 ± 0.012

Table 8.3: Total classification results, given as A_z , area under the ROC curve, for various combinations of features for the TB and the ID database. See text for detailed descriptions of the feature sets. Note that the working resolution is 700×700 pixels for the TB images and 1400×1400 pixels for the ID images, and the scale σ is given in pixels, so $\sigma = 1$ for the TB images corresponds to $\sigma = 2$ for the ID images.

database and the ID database. Table 8.4 lists the A_z values for each region for the TB database and the ID database, using the best performing feature sets.

Discussion

It is interesting to note that the overall performance of $A_z = 0.825$ on the TB database is much higher than the A_z values of the regions, as shown in Table 8.4, that do not

TB database							
region	A_z	region	A_z	region	A_z	region	A_z
0	0.73	12	0.68	24	0.77	36	0.83
1	0.76	13	0.71	25	0.79	37	0.57
2	0.76	14	0.63	26	0.55	38	0.67
3	0.68	15	0.57	27	0.60	39	0.74
4	0.67	16	0.63	28	0.73	40	0.66
5	0.63	17	0.76	29	0.69	41	0.78
6	0.64	18	0.63	30	0.71		
7	0.61	19	0.63	31	0.55		
8	0.58	20	0.64	32	0.60		
9	0.66	21	0.73	33	0.55		
10	0.54	22	0.80	34	0.67		
11	0.72	23	0.78	35	0.74		

ID database							
region	A_z	region	A_z	region	A_z	region	A_z
0	0.77	12	0.78	24	0.88	36	0.86
1	0.90	13	0.85	25	0.81	37	0.92
2	0.73	14	0.67	26	0.93	38	0.87
3	0.75	15	0.68	27	0.88	39	0.75
4	0.91	16	0.89	28	0.86	40	0.89
5	0.91	17	0.91	29	0.87	41	0.83
6	0.83	18	0.84	30	0.84		
7	0.83	19	0.82	31	0.66		
8	0.82	20	0.84	32	0.88		
9	0.75	21	0.80	33	0.86		
10	0.87	22	0.90	34	0.79		
11	0.86	23	0.87	35	0.91		

Table 8.4: Top: classification results of the TB database, per region, for the feature set consisting of difference features of 2nd moments with $\sigma = 2, 4$. Bottom: classification results of the ID database, per region, for the feature set consisting of 2nd, 3rd and 4th moments with $\sigma = 1, 2, 4$. Performance is given as A_z , area under the ROC curve. See Figure 8.2 for the region numbering scheme.

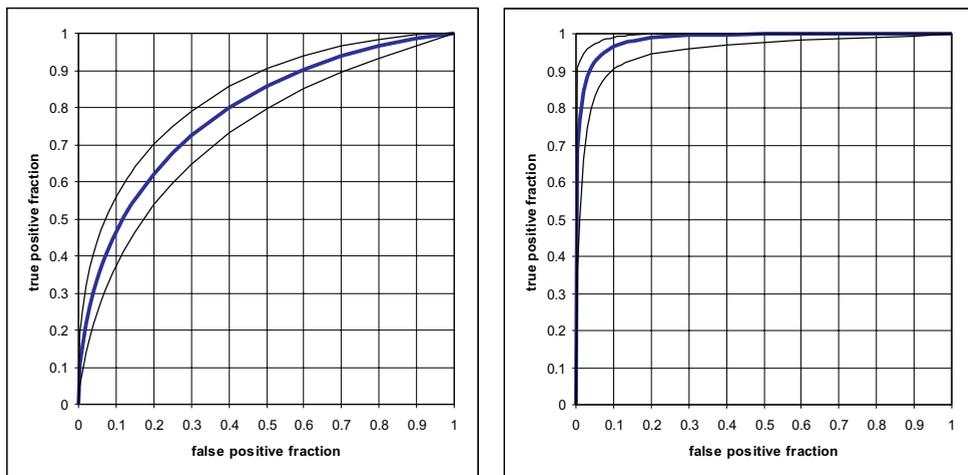


Figure 8.3: ROC curves for the best performing feature sets. The thin lines below and above the curve denote the asymmetric 95% confidence intervals. Left: the TB database, features set $\sigma = 2$, only 2nd moment and only difference features. The area under the curve is 0.820. Right: the ID database, feature set $\sigma = 1, 2, 4$, 2nd, 3rd and 4th moments. The area under the curve is 0.986.

exceed 0.70. Although other feature sets obtained occasional region scores up to over 0.80, this effect was consistent. Apparently, combining the scores improves performance. One reason for this effect could simply be statistical averaging. Abnormal images always contain at least three abnormal regions (due to overlapping regions and the definition of a region being abnormal if there is any overlap between the region and the abnormal area). Combining several estimates improves accuracy. Another reason might be that abnormal images are more likely to obtain higher abnormality scores for regions that are not outlined by the radiologist as abnormal compared to regions in normal images. In other words, radiologists may have outlined only the most abnormal regions in the abnormal images.

Let's consider how the components of the system affect the performance, that is, can better segmentation improve performance, is it possible to extract better features, is the classifier used for obtaining the score for each region optimal, is there room for improvement in combining the region scores into a single score for the complete image? The database itself is used in the classifier and is therefore also part of the system. So we should also discuss how the database affects performance.

Segmentation is most accurate for the vertical parts of the rib cage. Most failures occur in detecting the diaphragm and lung tops. If the lung top is detected at an incorrect height it is most often too low. This may lead to incorrect classification, because the areas over which texture features are collected do not correspond to the regions in the database. The same applies to the incorrectly detected diaphragm. Stomach gasses are difficult to deal with for Active Shape Models, which are based on the assumption of normal distribution of profiles. Apart from leading to incorrect

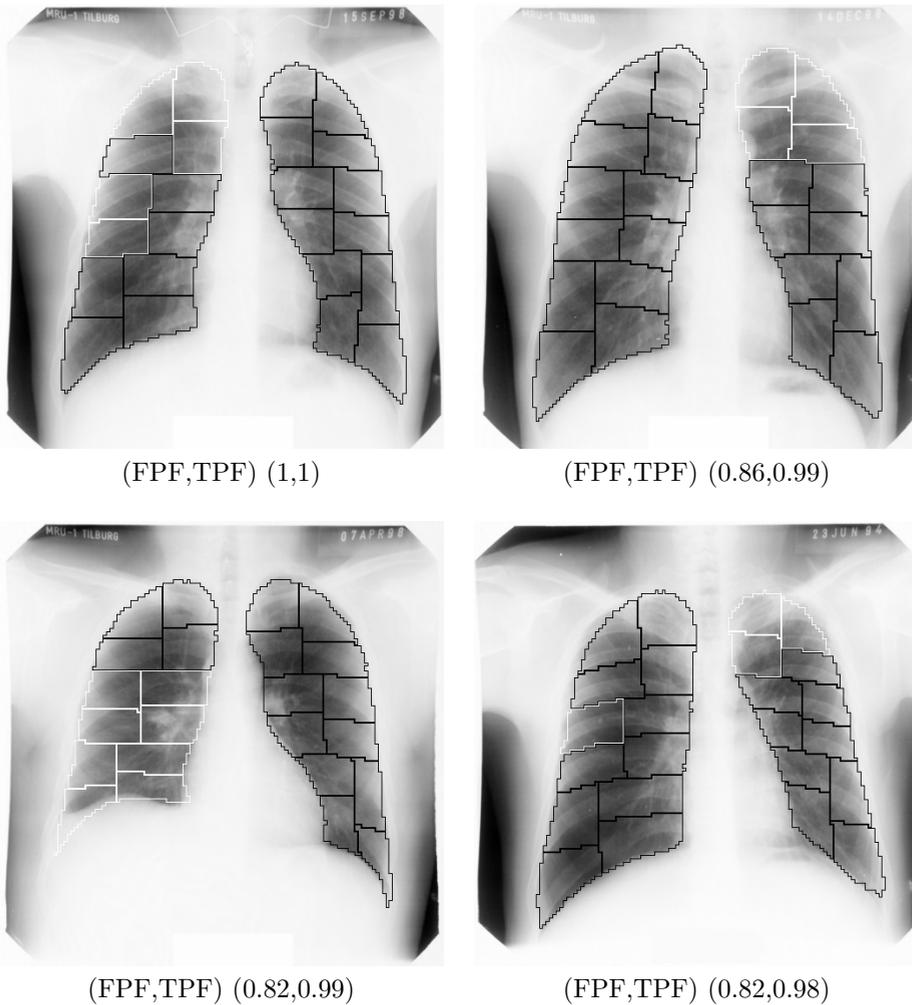
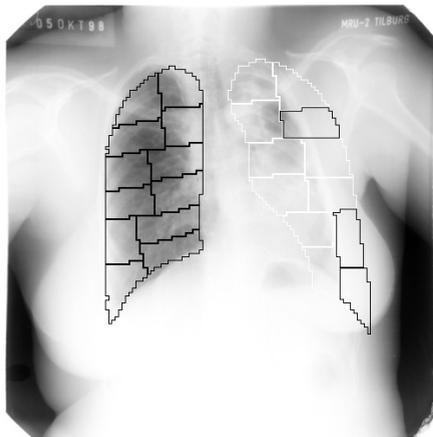
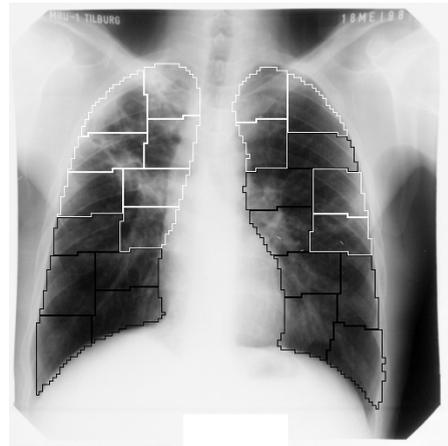


Figure 8.4: Of all abnormal cases in the TB database, the four images shown here were judged most normal according to the computerized scheme. These are the most difficult positive cases for the CAD scheme. Below each image its position on the ROC curve is indicated. The segmented regions are shown superimposed on the images. If a region is abnormal according to the radiologist, the outline of the segmentation is in white instead of black.

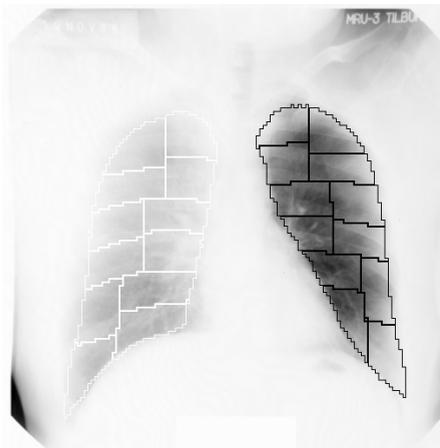
classification in case of segmentation errors, poor segmentation performance also leads to feature training sets that contain incorrect samples. Both effects indicate that the overall performance should increase if the segmentation scheme were improved. On the other hand, segmentation failures more often lead to images being classified as abnormal than normal and abnormal images are more likely to suffer from segmentation failures, as can be verified from Table 8.2. The overall effect of segmentation



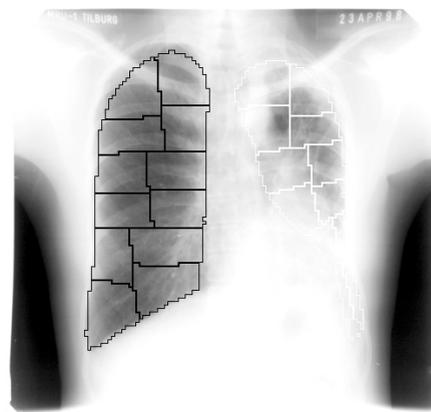
(FPF,TPF) (0,0.0068)



(FPF,TPF) (0,0.014)



(FPF,TPF) (0,0.020)

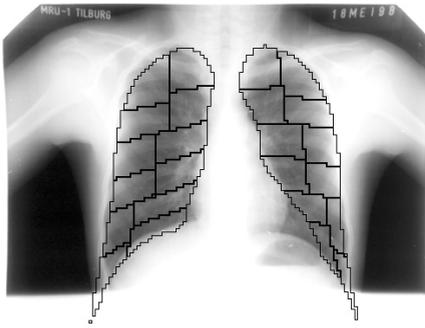


(FPF,TPF) (0,0.027)

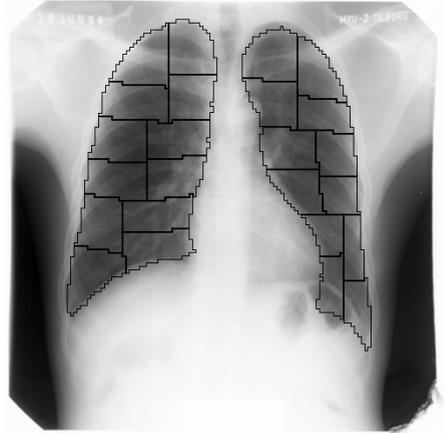
Figure 8.5: Of all abnormal cases in the TB database, the four images displayed here were the most abnormal according to the computerized scheme. These are the easiest positive cases for the CAD scheme, and indeed, the abnormalities are obvious. Below each image, its position on the ROC curve indicated. The segmented regions are shown superimposed on the images. If a region is abnormal according to the radiologist, the outline of the segmentation is in white instead of black.

errors on system performance could be tested by using manual segmentations.

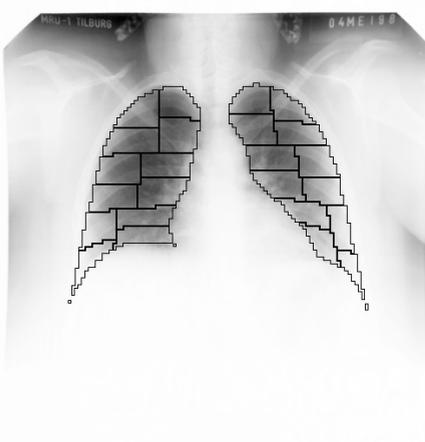
Table 8.3 shows that many different feature sets lead to approximately similar performance. In the case of different values of σ there is of course considerable correlation between the features. There are at least two general approaches to “more clever” features. The first is to ignore areas that appear to contain background structures, such as rib borders and large vessels. However, if these areas cannot be accurately de-



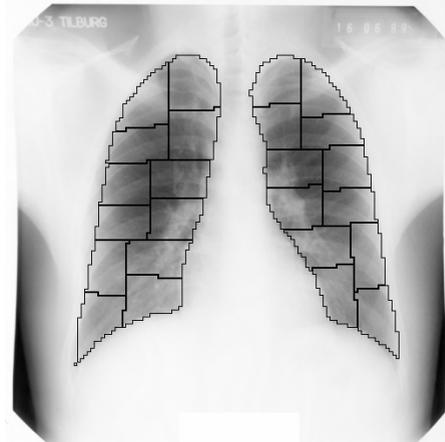
(FPF,TPF) (0.0041,0.21)



(FPF,TPF) (0.0083,0.25)



(FPF,TPF) (0.012,0.31)



(FPF,TPF) (0.017,0.35)

Figure 8.6: Of all normal cases in the TB database, the four images displayed here are the most abnormal according to the computerized scheme. These are the most difficult negative cases for the CAD scheme. Below each image, its position on the ROC curve indicated. The segmented regions are shown superimposed on the images.

tected, and/or abnormal structures are mistaken for normal background structures, the performance may actually deteriorate. Alternatively, pre-processing techniques to remove background structures could be employed, such as rib filtering [253] or contra-lateral subtraction [155].

With a total of 150 texture features available for the TB database, and 186 for the ID database, feature selection techniques could be employed, even separately for each region. This would, however, require the data sets to be split in two parts,

one for feature selection and one for testing. Since the number of samples is rather limited, we refrained from feature selection. Nevertheless, feature selection intuitively seems sensible. It is unlikely that the same features set would prove to be optimal for all regions. For instance, the difference features should be less effective in the lower medial region because of the heart shadow in the left lung field.

The classifier – the k NN classifier and the product rule with which the region scores are combined – is the least studied part of the system. It may well be possible that another classifier would yield slightly better results. However, in comparative studies [118] it has been shown that k NN classifiers often yield results comparable to or better than other classifiers, such as neural networks, and rarely show poor performance. Our attempts to improve the results by varying the parameters of the classification part of the systems, or by using other schemes to combine the region scores (averaging instead of multiplying, or using a k NN classifier) resulted in almost identical results.

For the TB database, abnormalities are highly concentrated in the upper lung fields. The number of abnormal samples is limited to values around 20 for lower and medial regions (Table 8.1). This is a very small number of samples, given that the variability in abnormal textures is high, to analyze with a texture analysis that involves far over a hundred features. It may be possible to improve the results by increasing the number of samples by adding samples from other regions. But this undermines a starting point of this work, namely that comparing local texture features with features from the same location in other radiographs should lead to a system that captures knowledge of the normal variation within each region and can distinguish this variation from the difference between normal and abnormal structures. A better solution is to expand the TB database. But note that the current database, with only about 20 abnormal samples for the lower and medial lung regions, is obtained from about 25.000 screenings. This emphasizes the need to establish huge databases in order to obtain progress in computer-aided diagnosis.

Our results on the ID database are comparable to, or even slightly better than, those reported by Ishida [111] and Katsuragawa [128] and co-workers. They used a combined scheme for the detection of interstitial disease evaluated on the same database. There are however a few distinctive differences between their approaches and the scheme presented here. First of all, they use a smaller part of the images for analysis, consisting of selected ROIs in the lateral lung fields, excluding the lung tops. Therefore their method is not directly applicable to the TB database since most abnormal findings in this database are located in the lung tops. Second, they perform a single classification for all these regions and not a specific one for each region. Third, they compute only a distance from the normal case, and do not use the statistics of abnormal regions directly in the classification. Fourth, these studies use a smaller set of specifically selected features. With these features, they are able to divide the abnormal findings in three classes (reticular, nodular and honeycomb patterns).

If our system were used to automatically detect normal cases in TB screening, so that these images would not have to be read by a human expert anymore, the current results indicate that about 15% of abnormal cases would be incorrectly classified as normal if 50% of all images were classified as normal (operating point on the ROC curve of $(FPF, TPF) = (0.50, 0.86)$). This is probably not good enough for application

in practice and one has to keep in mind that these results only apply to “textural abnormalities” and images without clothing artifacts.

Assuming the availability of far more data, which would allow feature selection methods for each region, and an “improved” classifier, is it realistic to assume that the results for the TB database would approach those for the ID database? We do not believe this to be the case, although performance improvements may be possible. The normal images in Figure 8.6 that are erroneously judged to be abnormal might be correctly classified when more statistical data are available. But there seems to be a structural difference between both databases. Although they both contain clearly abnormal images that are correctly estimated to be very abnormal, such as the ones shown in Figure 8.5, the TB database has cases with extremely subtle abnormalities, limited to small areas of the image. The abnormal areas in the images in the ID database are in almost all cases larger than one third of a complete lung field. Therefore substantially more statistical information about the appearance of abnormalities is available for the ID database, although the total number of abnormal images is smaller. Another difference is the resolution of the images. There is some information in the finest details of the ID database images that is not available for the TB database images. Subsampling the ID database images to 700×700 pixels, the same as the TB database images, and running the texture analysis yielded $A_z = 0.960$ for the best performing feature set; lower than the $A_z = 0.986$ but still much better than the TB database. Therefore we believe the difference in the nature and subtlety of the abnormalities found in both databases is the most important reason why performance is different. Consider the abnormal images in Figure 8.4 that were classified as highly normal by the computer. These are examples of images that would be “missed” in a computerized screening as explained above. The two images on the top row and the bottom-right image contain highly subtle abnormalities. The image at the bottom-left contains a clear abnormality next to the right hilum, but apparently the normal variation between hilum regions is so great and the number of abnormal samples in this region so small (remember that enlarged hila were excluded from the TB database in this study) that the texture classification scheme does not detect it. Note also that A_z is not high for these regions, so that they are not weighed heavily when combining the results from separate regions into the overall abnormality score.

Conclusion

We have presented a fully automatic scheme for texture analysis of lung fields in chest radiographs. The method is based on texture analysis on *local* regions in the image, which are segmented automatically. Features are extracted from histograms of the responses of a multi-scale filter bank. Each region is processed independently with a different k -nearest neighbor classifier. By accumulating local evidence an overall abnormality score for the complete image can be computed. Experiments were performed on a database containing chest radiographs with interstitial disease with excellent results. On a large database from a mass chest screening against TB with images with textural abnormalities, containing many subtle cases, the results were fairly accurate. Although not all classes of abnormal findings from this TB

screening were used in the experiments, the results suggest that this method may be helpful to radiologists for reading mass chest screening images.

Chapter 9

Summary and General Discussion

ABOUT 40% OF ALL RADIOGRAPHS are chest radiographs. The chest exam is so important because it can identify a wide variety of abnormal conditions in a patient. Yet even experienced radiologists often have great difficulty in interpreting the signs on a chest radiograph. This explains the wide interest in the subject of this thesis, computer-aided diagnosis in chest radiography. In this final chapter we summarize the obtained results and discuss possible future directions for computer-aided diagnosis in mass chest screening against tuberculosis and in chest radiography in general.

Summary

The literature on computer analysis of chest radiographs, which is reviewed in **Chapter 2**, encompasses a large range of topics. Segmentation of the lung fields, the rib cage and the heart has received much attention. Many studies attempt to characterize the textural appearance of the lung fields, so as to detect interstitial disease in general or for specific applications, notably detection of pneumoconiosis. Other topics are the development of subtraction techniques to suppress normal structures and thereby enhance abnormalities; detection algorithms for catheters; and methods to detect pneumothoraces. The subject that has received most attention is the detection and characterization of lung nodules. At the end of **Chapter 2** some directions for future research are given. We will return to these recommendations in the next section and discuss them in the light of the results of this thesis, which has focused on statistical segmentation techniques and texture analysis methods, which have been tested on data from a mass chest screening program against tuberculosis.

Chapter 3 describes nine algorithms for the automatic delineation of lung fields in chest radiographs. The algorithms are based on different techniques: matching, pixel classification based on several combinations of features, and a rule-based scheme. The latter uses oriented edges and ridges at optimal scales, to detect the different parts of the lung contours. The rule-based scheme and the pixel classifier approach are combined in a hybrid scheme. This hybrid scheme yields the best performance

in experiments in which all systems are tested on 115 images. Comparisons are made with results available from the literature. The hybrid scheme has a robust performance: for all images in the test set the accuracy is above 94%. The average accuracy of the scheme is 0.969 ± 0.008 , which is close to the inter-observer variability of 0.984 ± 0.005 .

In **Chapter 4** several applications of locally orderless images (LOIs), introduced by Koenderink and Van Doorn [141], are proposed. One can think of these entities as the proper way to make a histogram, in the same sense as scale-space theory has shown that, under very general conditions, convolution with a Gaussian kernel is the proper way to reduce the spatial resolution of an image. LOIs are a family of three intertwined scale-spaces. The image is represented, at an inner scale σ , as a local histogram with bin-width β , at each location, with a Gaussian-shape region of interest of extent α . Applications are presented for a range of image processing tasks, including new non-linear diffusion schemes, adaptive histogram equalization and variations, several methods for noise and scratch removal, texture rendering, and texture segmentation.

Apart from these applications, LOIs can also be extended to a general framework for texture feature extraction. This is the topic of **Chapter 5**. LOIs are generalized by considering local histograms of a collection of filtered versions of the image, and by extending them to joint probability distributions. From these constructions, texture features can be derived. It is shown that this framework includes two established texture classification methods, namely filter bank methods and co-occurrence matrices. Several extensions to these classical methods that have been proposed in the literature are also part of the LOI texture framework. Because all scale parameters are stated explicitly, various multi-resolution feature sets can be extracted in a systematic way, and it becomes straightforward to perform new types of multi-resolution texture analysis that have not been considered before. We propose to use a collection of oriented Gaussian derivatives as filters because they describe the local jet, a Taylor expansion of the image. Experiments are described that show that *each* multi-resolution approach improves the performance of texture classification schemes. The best result is obtained with a multi-resolution filter bank approach in which both σ and α are varied.

In **Chapter 6** a new version of the active shape models (ASMs) segmentation scheme from Cootes and Taylor [49, 48] is presented. The ASM scheme is a general model-based method: trained with a number of examples in which the objects to be segmented are indicated with a fixed number of landmarks, it can be applied, in principal, to *any* segmentation task. In that regard the method is different from the hand-crafted methods for lung field segmentation presented in Chapter 3. In its original formulation, the ASMs method uses normalized first order derivatives profiles, sampled at multiple scales perpendicular to each landmark point, to steer the iterative segmentation process. The new method presented in this chapter uses *optimal local* features instead. The features are local texture descriptors derived from the framework of Chapter 5. A non-linear k NN-classifier is used to find optimal displacements for landmarks, instead of the linear Mahalanobis distance used in the original ASM scheme. The optimal features are automatically selected from the training images.

The method is tested on five different segmentation tasks. One task uses simulated data in which a house-like object with varying shape is filled with a specific texture and the background is filled with an other texture. It is demonstrated that the original scheme cannot detect objects whose boundaries are formed by a transition of two textures, but the method with optimal features can deal with such boundaries. The four other tasks concern segmentation of medical images: finding the right and left lung fields in chest radiographs and the cerebellum and corpus callosum in slices from MRI brain images. For each task, the ASM scheme with optimal features produces significantly better results ($p < 0.001$) than the original scheme.

Chapter 7 is about rib segmentation. An automatic method for segmenting the posterior ribs in frontal chest radiographs is presented, based on a statistical shape model for the complete rib cage. This is different from previous work, in which candidate rib borders were analyzed one-by-one and merged with rules into a complete rib cage. The advantage of using a complete rib cage model is that correlations between the ribs are taken explicitly into account. The lower ribs are often hardly visible in chest radiographs, yet human observers can still discern them, because they can infer their location from the location of the higher ribs that are visible more clearly. In a similar way, the rib segmentation method presented here combines image based evidence of ribs that are clearly visible, with knowledge of the shape correlations between ribs. Therefore rib borders that are difficult to discern are put at (at least approximately) the correct position. Each posterior rib is modeled by two equidistant parabolas. This requires four parameters per rib so the full rib cage, from rib 2 through rib 10, contains 72 parameters. This number is reduced with principal component analysis; it is demonstrated that 10 parameters explain over 98% of the variability in a training set of 35 chest radiographs. The rib cage is fitted with Powell's direction set method for optimizing the model parameters, with a fit measure that accumulates the edge strength perpendicular to the rib borders at a fixed scale. It turns out that there exists a range of optimal scales and a fixed scale in that range is selected. The method is tested on 35 images with a resolution of 512 by 512 pixels. Rib borders are located with an accuracy of 3 pixels on average.

In **Chapter 8** the two main threads of this thesis, segmentation methods and texture analysis, are combined into a fully automatic method to detect abnormalities in chest radiographs and to compute a single overall estimate of the probability that the image is abnormal. The method is designed for finding abnormal signs of a diffuse, textural nature, such as they are encountered in mass chest screening against tuberculosis. The scheme starts with automatic segmentation of the lung fields, using active shape models. The segmentation is used to subdivide the lung fields into overlapping regions of various sizes. Texture features are extracted from each region, using the moments of responses to a multi-scale filter bank. Additional difference features are obtained by subtracting feature vectors from corresponding regions in the left and right lung fields. A separate training set is constructed for each region. All regions are classified by voting among the k nearest neighbors, with a leave-one-out strategy. Next, the classification results of each region are combined, using a weighted multiplier in which regions with higher classification reliability weigh more heavily. This produces an abnormality score for each image. The method is evaluated

on two databases. The first database was collected from a TB mass chest screening program, from which 147 images with textural abnormalities and 241 normal images were selected. Although this database contains many subtle abnormalities, the area under the ROC curve is 0.820 and the classification has a sensitivity of 0.86 at a specificity of 0.50. This means that if 50% of the images in a tuberculosis screening program would be automatically classified as normal by this computer scheme and not read by radiologists, 14% of all abnormal cases would be missed. Note however that only images with textural abnormalities have been included in this study, so this number cannot be directly generalized to a true screening (although the majority of the abnormal cases encountered in TB screening falls into this category of textural abnormalities). The second database on which the method is tested consisted of 100 normal images and 100 abnormal images with interstitial disease obtained from the University of Chicago Hospitals. For this database the results are excellent: a sensitivity of 0.97 at a specificity of 0.90 and an area under the ROC curve of 0.986.

General discussion

This PhD project was initiated in cooperation with Nucletron, a Dutch company that manufactures the Oldelft Digidelca-M, a digital chest unit for the mass chest screening market. The goal was to design a system that can automatically classify images from a TB screening as very likely normal or possibly abnormal. The method presented in Chapter 8 could be considered a first attempt at the design of such a system. It is a *proof of principle* that automatic segmentation and general texture analysis can detect a large part of the abnormal cases in a TB screening, with only a limited number of false positives, *i.e.* missed cases (14% at a false positive rate of 50%). This result is probably not good enough for commercial application, although this decision ultimately has to be made by those who organize the screening. Any screening suffers from missed cases and a discussion on the proper way to measure the effectiveness of strategies to fight the TB epidemic is far outside the scope of this thesis. Besides their possible use in mass chest screening, the techniques developed in this thesis can also be used in CAD systems that analyze clinical chest radiographs as they are obtained routinely in a general hospital. In this general discussion we consider how the obtained results could be improved. There is some overlap with the recommendations for further research given at the end of Chapter 2.

Two research areas have been central in this work: texture analysis and segmentation. Histograms feature prominently in most texture analysis schemes, so the framework of locally orderless images (scale-space histograms) turned out to be ideally suited to describe a general texture analysis method (Chapters 4 and 5). This method encompasses filter bank methods, which are currently state-of-the-art in texture analysis. The framework from Chapter 5 sheds new light on the use of multi-resolution approaches.

Texture classification has always been knowledge-based, because it relies on classifiers, trained with examples. Segmentation methods have often been constructed according to a step-by-step design, as we did in Chapter 3 in a rule-based system for segmenting the lung fields. But there appears to be a paradigm shift towards the

use of statistical evidence to steer segmentation methods. Some authors [20] divide the computer vision community in supporters of the bottom-up versus the top-down approach. The bottom-up community develops generic tools to detect structures such as edges, corner points, and so on, in images. They have developed general-purpose segmentation techniques. The top-down community has abandoned the search for the Holy Grail of general-purpose computer vision systems, and focuses on solving problems in specific domains by incorporating prior knowledge about the task to be solved into the algorithms. For segmentation, general and knowledge-based methods such as active shape models are available. Our contribution to this field, applied to segmentation problems in chest radiography, can be found in Chapter 6 on optimal features for active shape models and in Chapter 7 on statistical rib cage segmentation. There is still much room for improvement in such methods; they use effective, but rather simple techniques as principal component analysis and usually process the “raw pixels”. Ideally, the results of the bottom-up researchers, who have established robust methods to describe local structure in images, should be incorporated into knowledge-based algorithms. The system put forward in Chapter 6 could be considered a first attempt to do this. From the results of Chapter 7 it will be obvious that further improvements are needed: the problem of segmenting the rib cage in chest radiographs is still far from solved (see also Chapter 2). An accurate segmentation of the rib cage, if available, could be used in subsequent processing, something that has been omitted in Chapter 8.

Several issues that are important in the field of computer analysis of chest radiographs have not been considered in this thesis.

First, we have not paid specific attention to the use of sophisticated classifiers. Our experience suggests that the different classifiers available, see *e.g.* [118], give comparable results on our data, but this experience is not based on extensive research. It is possible that one can design a new or take an existing classifier – probably in combination with a mechanism to select or weigh features – that significantly improves performance for CAD problems in chest radiography, especially if much more data are available, a point to which we will return shortly.

Second, sophisticated subtraction algorithms are likely to become standard in pre-processing for CAD in chest radiography. It has been shown that temporal subtraction can be very useful to detect interval change. Contralateral subtraction may identify faint lesions. And statistical models of the appearance of normal chest radiographs may be fitted to input images and subtracted to enhance possibly abnormal structures. An advanced CAD system will likely produce a number of different input images for each radiograph to be analyzed. In many digital chest units it is technically feasible to make two photographs with different energies (kV) at the same time and to subtract these in order to obtain a dual energy subtraction image in which the bony structures are virtually invisible. Such images have never been used routinely in clinical practice, but they can be extremely useful for CAD. An example is shown in Figure 9.1.

Third, nodule detection is the most studied problem in CAD for chest radiographs. Although most abnormalities in TB screening are of a textural nature, techniques that can detect abnormalities that take the form of dense objects in the lung fields should be an integral part of any advanced system that analyzes chest radiographs from TB screening.



Figure 9.1: An example of the usefulness of dual energy subtraction radiography. On the left a part of a conventional chest radiograph is shown. In the middle the same patient is shown, photographed at the same time with a different kV that shows only bony structures. This image could be used to detect the ribs and other bones. On the right the difference between the left and the middle image is shown. In this dual energy subtraction image the abnormality is much more clearly visible. Images courtesy of the University of Chicago.

The most important step to improve current state-of-the-art in CAD for chest radiography is to use much larger databases. For the research described in this thesis digital images were not available and a limited set of small-size films had to be scanned manually. With current hardware – chest units and general computers – and the advent of digital radiology departments, it should be possible to use much larger databases of digital images in future research. In older articles, researchers evaluated their techniques on only a handful of images (see Chapter 2). Today, evaluations on a few hundred cases are becoming standard practice. These numbers will increase in the future, which will make it possible to divide problems into sub-domains. Our TB database was simply too small to do this. The first CAD applications in chest radiography will likely work as background processes in a clinical chest workstation, much like you use a spell-checker to go over typed text in a word-processor. In a chest workstation, clinical information such as the age and sex of the subject, the type of acquisition and the reason for the examination, is always available. The databases and specific CAD methods should be selected based on this information. The scientific community has acknowledged by and large that general computer vision systems are an illusion and that we have to train systems to work in very specific domains. With huge databases, such specific subsystems can be built.

In contrast to this perspective, the system outlined in Chapter 8 is a “shoot-at-everything” method: it uses the same texture features for all abnormal findings, without correcting for the variations in normal anatomy, and uses only a rough segmentation of the lung fields. The abnormalities encountered in TB screening can be subdivided in many ways, and in theory, with enough data, the method outlined in Chapter 8 could be optimized for every different type of abnormality. This could lead to a system in which each input image has to pass a series of tests for different abnormalities, each optimized for the sex and age group of the patient and the peculiarities

of the image acquisition system and the screening population.

The same applies to segmentation problems. The age of the patient has a large impact on the appearance of the anatomy. And there are specific classes of images that should be treated separately. For example, the stomach bubble below the left lung is invisible in some images, but can be as large as the lung field in other cases. Segmentation algorithms should detect this and choose a specific strategy based on the result of this detection step.

We conclude with a summary of points that are important to improve CAD systems for TB screening and chest radiography in general:

- Use much larger databases, split the problem up into sub-problems, and use specific databases for each subproblem; use different databases for different patient groups *e.g.* based on age and sex.
- Do more intelligent segmentation, based on knowledge-based segmentation techniques. Use the segmentation of the lung fields, the heart, the posterior and anterior ribs, the clavicles and the major vessels in subsequent processing.
- Use techniques to subtract normal structures and enhance abnormalities. In practice this pre-processing will produce a range of input images. If dual energy imaging is available, the system should take this data into account as well.
- More sophisticated CAD systems will probably use more sophisticated pattern classifiers than the ones used in this thesis.

Bibliography

- [1] K. Abe, K. Doi, H. MacMahon, M.L. Giger, H. Jia, X. Chen, A. Kano, T. Yanagisawa, “Computer-aided diagnosis in chest radiography, preliminary experience”, *Investigative Radiology*, vol. 28, no. 11, pp. 987–993, 1993. [21](#)
- [2] R. Adams & L. Bischof, “Seeded region growing”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 6, pp. 641–647, 1994. [72](#)
- [3] A.A. Amini, T.E. Weymouth, R.C. Jain, “Using dynamic programming for solving variational problems in vision”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 9, pp. 855–867, 1990. [46](#)
- [4] S.G. Armato, M.L. Giger, K. Ashizawa, H. MacMahon, “Automated lung segmentation in digital lateral chest radiographs”, *Medical Physics*, vol. 24, no. 8, pp. 1507–1520, 1998. [22](#), [22](#), [23](#)
- [5] S.G. Armato, M.L. Giger, H. MacMahon, “Computerized detection of abnormal asymmetry in digital chest radiographs”, *Medical Physics*, vol. 21, no. 11, pp. 1761–1768, 1994. [26](#)
- [6] S.G. Armato, M.L. Giger, H. MacMahon, “Automated lung segmentation in digitized postero-anterior chest radiographs”, *Academic Radiology*, vol. 4, pp. 245–255, 1998. [22](#), [23](#), [137](#)
- [7] S.G. Armato, M.L. Giger, H. MacMahon, “Computerized delineation and analysis of costophrenic angles in digital chest radiographs”, *Academic Radiology*, vol. 5, pp. 329–335, 1998. [22](#), [22](#), [23](#)
- [8] S. Arya & D.M. Mount, “Approximate nearest neighbor queries in fixed dimensions”, in *Proceedings of the 4th ACM-SIAM Symposium on Discrete Algorithms*, pp. 271–280, 1993. [58](#), [58](#), [85](#), [105](#), [140](#)
- [9] N. Asada, K. Doi, H. MacMahon, S. Montner, M.L. Giger, C. Abe, Y. Wu, “Potential usefulness of an artificial neural network for differential diagnosis of interstitial lung diseases: pilot study”, *Radiology*, vol. 177, no. 13, pp. 857–860, 1990. [20](#)
- [10] Y. Asai, Y. Tanabe, Y. Ozaki, H. Kubota, M. Matsumoto, H. Kanamori, “Optimum tube voltage for chest radiographs obtained by psychophysical analysis”, *Medical Physics*, vol. 25, no. 11, pp. 2170–2175, 1998. [20](#), [40](#)
- [11] K. Ashizawa, T. Ishida, H. MacMahon, C.J. Vyborny, S. Katsuragawa, K. Doi, “Artificial neural networks in chest radiography”, *Academic Radiology*, vol. 6, pp. 2–9, 1999. [20](#)
- [12] K. Ashizawa, H. MacMahon, T. Ishida, K. Nakamura, C.J. Vyborny, S. Katsuragawa, K. Doi, “Effect of an artificial neural network on radiologists’ performance in the differential diagnosis of interstitial lung disease using chest radiographs”, *American Journal of Roentgenology*, vol. 172, pp. 1311–1315, 1999. [20](#)
- [13] D.H. Ballard, *Hierarchical recognition of tumors in chest radiographs*, Birkhauser Verlag, New York, 1976. [29](#), [30](#), [30](#), [30](#), [30](#), [31](#)
- [14] D.H. Ballard, “Model-directed detection of ribs in chest radiographs”, in *Proceedings of the 4th International Joint Conference on Pattern Recognition*, pp. 907–910, 1978. [25](#), [26](#), [118](#)
- [15] D.H. Ballard & J. Sklansky, “A ladder-structured decision tree for recognizing tumors in chest radiographs”, *IEEE Transactions on Computers*, vol. 20, pp. 503–513, 1976. [29](#), [30](#), [30](#), [31](#)

- [16] H.C. Becker, W.J. Nettleton, P.H. Meyers, J.W. Sweeney, C.M. Nice Jr., "Digital computer determination of a medical diagnostic index directly from chest X-ray images", *IEEE Transactions on Biomedical Engineering*, vol. BME-11, pp. 67–72, 1964. [19](#), [28](#)
- [17] G. Behiels, D. Vandermeulen, F. Maes, P. Suetens, P. Dewaele, "Active shape model-based segmentation of digital X-ray images", in *MICCAI '99*, no. 1679 in Lecture Notes in Computer Science, Springer, Berlin, pp. 128–137, 1999. [97](#), [101](#), [101](#), [115](#)
- [18] R.E. Bellman, *Applied Dynamic Programming*, Princeton University Press, 1962. [46](#)
- [19] F. Bergholm, "Edge focusing", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 6, pp. 726–741, 1987. [56](#)
- [20] A. Blake & M. Isard, *Active contours*, Springer, 1998. [96](#), [159](#)
- [21] J-P. Bolet, A.R. Cowen, J. Launders, A.G. Davies, R.F. Bury, "Progress with an 'all-wavelet' approach to image enhancement and de-noising of direct digital thorax radiographic images", in *Proceedings of the IPA'97*, pp. 244–248, 1997. [21](#)
- [22] F.L. Bookstein, "Landmark methods for forms without landmarks: morphometrics of group differences", *Medical Image Analysis*, vol. 1, no. 3, pp. 225–244, 1997. [96](#)
- [23] J.M. Boone & J. Duryea, "Filter wheel equalization for chest radiography: a computer simulation", *Medical Physics*, vol. 22, no. 7, pp. 1029–1037, 1995. [21](#)
- [24] A.C. Bovik, M. Clark, W.S. Geisler, "Multichannel texture analysis using localized spatial filters", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 55–73, 1990. [76](#), [134](#)
- [25] M. Brejl & M. Sonka, "Object localization and border detection criteria design in edge-based image segmentation: automated learning from examples", *IEEE Transactions on Medical Imaging*, 2000. To appear. [96](#), [101](#), [106](#)
- [26] P. Brodatz, *Textures*, Dover, New York, 1966. [70](#), [84](#), [106](#)
- [27] L.G. Brown, "A survey of image registration techniques", *ACM Computing Surveys*, vol. 24, no. 4, pp. 325–376, 1992. [41](#)
- [28] M.S. Brown, L.S. Wilson, B.D. Doust, R.W. Gill, C. Sun, "Knowledge-based method for segmentation and analysis of lung boundaries in chest X-ray images", *Computerized Medical Imaging and Graphics*, vol. 22, pp. 463–477, 1998. [23](#), [40](#), [137](#)
- [29] R.E. Bunge & C.L. Herman, "Use of diagnostic imaging procedures: a nationwide hospital study", *Radiology*, vol. 163, no. 2, pp. 569–573, 1987. [15](#), [19](#)
- [30] D. Burr, C. Morrone, D. Spinelli, "Evidence of edge and bar detectors in human vision", *Vision Research*, vol. 29, pp. 419–431, 1989. [76](#)
- [31] J. Canny, "A computational approach to edge detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986. [43](#)
- [32] F.M. Carrascal, J.M. Carreira, M. Souto, P.G. Tahoces, L. Gomez, J.J. Vidal, "Automatic calculation of total lung capacity from automatically traced lung boundaries in postero-anterior and lateral digital chest radiographs", *Medical Physics*, vol. 25, no. 7, pp. 1118–1131, 1998. [22](#), [22](#), [23](#), [29](#), [39](#)
- [33] M.J. Carreira, D. Cabello, M.G. Penedo, A. Mosquera, "Computer-aided diagnoses: automatic detection of lung nodules", *Medical Physics*, vol. 25, no. 10, pp. 1998–2006, 1998. [32](#)
- [34] M. Casaldi, G. Russo, G. Scarano, P. Talone, "Automatic detection of lung nodules: application to radiograms lossy coding", in *Proceedings SPIE*, vol. 3338, pp. 1365–1376, 1998. [32](#)
- [35] D.C. Chan & W.R. Wu, "Image contrast enhancement based on a histogram transformation of local standard deviation", *IEEE Transactions on Medical Imaging*, vol. 17, no. 4, pp. 518–531, 1998. [68](#)
- [36] X. Chen, K. Doi, S. Katsuragawa, H. MacMahon, "Automated selection of regions of interest for quantitative analysis of lung textures in digital chest radiographs", *Medical Physics*, vol. 20, no. 4, pp. 975–982, 1993. [26](#), [34](#), [134](#), [134](#)

- [37] Y.P. Chien & K.S. Fu, “A decision function method for boundary detection”, *Computer Graphics and Image Processing*, vol. 3, pp. 125–140, 1974. [23](#)
- [38] Y.P. Chien & K.S. Fu, “Recognition of X-ray picture patterns”, *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-4, no. 2, pp. 145–156, 1974. [23](#)
- [39] H.G. Chotas & C.E. Ravin, “Chest radiography: estimated lung volume and projected area obscured by the heart, mediastinum, and diaphragm”, *Radiology*, vol. 193, no. 2, pp. 403–404, 1994. [41](#)
- [40] M.L. Cocklin, A.R. Gourlay, P.H. Jackson, G. Kaye, I.H. Kerr, Lams P., “Digital processing of chest radiographs”, *Image and Vision Computing*, vol. 1, no. 2, pp. 67–78, 1983. [20](#)
- [41] R.W. Connors & C.A. Harlow, “Equal probability quantizing and texture analysis of radiographic images”, *Computer Graphics and Image Processing*, vol. 8, pp. 447–463, 1978. [21](#)
- [42] R.W. Connors & C.A. Harlow, “A theoretical comparison of texture algorithms”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-2, no. 3, pp. 204–222, 1980. [76](#)
- [43] R.W. Connors & C.A. Harlow, “Toward a structural textural analyzer based on statistical methods”, *Computer Graphics and Image Processing*, vol. 12, pp. 224–256, 1980. [76](#)
- [44] R.W. Connors, C.A. Harlow, S.J. Dwyer III, “Radiographic image analysis: past and present”, *??*, pp. 1152–1168, 1982. [19](#), [21](#)
- [45] L.T. Cook, G.G. Cox, M.F. Insana, M.A. McFadden, T.J. Hall, R.S. Gaborski, F.Y.M. Lure, “Comparison of a cathode-ray-tube and film for display of computed radiographic images”, *Medical Physics*, vol. 25, no. 7, pp. 1132–1138, 1998. [20](#)
- [46] L.T. Cook, M.F. Insana, M.A. McFadden, T.J. Hall, G.G. Cox, “Contrast-detail analysis of image degradation due to lossy compression”, *Medical Physics*, vol. 22, no. 6, pp. 715–721, 1995. [20](#)
- [47] T.F. Cootes, A. Hill, C.J. Taylor, J. Haslam, “The use of active shape models for locating structures in medical images”, *Image and Vision Computing*, vol. 12, no. 6, pp. 355–366, 1994. [97](#), [101](#)
- [48] T Cootes & C.J. Taylor, “Statistical models of appearance for computer vision”, Tech. Rep., Wolfson Image Analysis Unit, University of Manchester, 1999. [95](#), [96](#), [97](#), [97](#), [120](#), [138](#), [156](#)
- [49] T.F. Cootes, C.J. Taylor, D. Cooper, J. Graham, “Active shape models – their training and application”, *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995. [24](#), [95](#), [96](#), [96](#), [120](#), [138](#), [156](#)
- [50] G.G. Cox, L.T. Cook, M.F. Insana, M.A. McFadden, T.J. Hall, L.A. Harrison, D.A. Eckard, N.L. Martin, “The effects of lossy compression on the detection of subtle pulmonary nodules”, *Medical Physics*, vol. 23, no. 1, pp. 127–132, 1996. [20](#)
- [51] G.S. Cox, F.J. Hoare, G. de Jager, “Experiments in lung cancer nodule detection using texture analysis and neural network classifiers”, in *Third South African Workshop on Pattern Recognition*, 1992. [31](#)
- [52] G.R. Cross & A.K. Jain, “Markov random field texture models”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 5, no. 1, pp. 25–38, 1983. [77](#)
- [53] K.J. Dana, B. van Ginneken, S.K. Nayar, J.J. Koenderink, “Reflectance and texture of real-world surfaces”, *ACM Transactions on Graphics*, vol. 18, no. 1, pp. 1–34, 1999. [70](#)
- [54] L.S. Davis, S.A. Johns, J.K. Aggerwal, “Texture analysis using generalized co-occurrence matrices”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 1, no. 3, pp. 215–259, 1979. [76](#), [80](#)
- [55] P. de Souza, “Automatic rib detection in chest radiographs”, *Computer Vision, Graphics, and Image Processing*, vol. 23, pp. 129–161, 1983. [24](#), [25](#), [118](#)
- [56] M.C. Difazio, H. MacMahon, X-W. Xu, P. Tsai, J. Shiraishi, S.G. Armato III, K. Doi, “Digital chest radiography: effect of temporal subtraction images on detection accuracy”, *Radiology*, vol. 202, pp. 447–452, 1997. [22](#)

- [57] K. Doi, “Perspectives on digital image analysis in medical imaging: needs for a new science concerning technical understanding of the contents of medical images”, *ICRU News*, vol. 1, pp. 10–14, June 1996. [21](#)
- [58] K. Doi, “Computer-aided diagnosis and its potential impact on diagnostic radiology”, in *Computer-aided diagnosis in medical imaging*, K. Doi, H. MacMahon, M.L. Giger, K.R. Hoffmann (eds.), Elsevier, 1999. [21](#)
- [59] K. Doi, M.L. Giger, R.M. Nishikawa, K.R. Hoffmann, H. MacMahon, R.A. Schmidt, “Potential usefulness of digital imaging in clinical diagnostic radiology: computer-aided diagnosis”, *Journal of Digital Imaging*, vol. 8, no. 1, pp. 2–7, 1995. [21](#)
- [60] K. Doi, M.L. Giger, R.M. Nishikawa, K.R. Hoffmann, H. MacMahon, R.A. Schmidt, K-G. Chua, “Digital radiography: a useful clinical tool for computer-aided diagnosis by quantitative analysis of radiographic images”, *Acta Radiologica*, vol. 34, pp. 426–439, 1993. [21](#)
- [61] K. Doi, M.L. Giger, R.M. Nishikawa, K.R. Hoffmann, H. MacMahon, R. Schmidt, C.E. Metz, “Recent progress in development of computer-aided diagnostic (CAD) schemes in radiology”, *Medical Imaging Technology*, vol. 13, no. 6, pp. 822–835, 1995. [21](#)
- [62] K. Doi, H. MacMahon, S. Katsuragawa, R.M. Nishikawa, Y. Jiang, “Computer-aided diagnosis in radiology: potential and pitfalls”, *European Journal of Radiology*, vol. 31, pp. 97–109, 1997. [21](#)
- [63] I. Dryden & K.V. Mardia, *The statistical analysis of shape*, Wiley, London, 1998. [96](#), [97](#), [120](#)
- [64] J.S. Duncan & N. Ayache, “Medical image analysis: progress over two decades and the challenges ahead”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 85–106, 2000. [19](#), [36](#), [36](#), [36](#)
- [65] J. Duryea & J.M. Boone, “A fully automatic algorithm for the segmentation of lung fields in digital chest radiographic images”, *Medical Physics*, vol. 22, no. 2, pp. 183–191, 1995. [22](#), [23](#), [39](#), [53](#), [53](#), [53](#), [137](#)
- [66] R.L. Engle Jr., “Attempts to use computers as diagnostic aids in medical decision making: a thirty-year experience”, *Perspectives in Biology and Medicine*, vol. 35, no. 2, pp. 207–217, 1992. [19](#)
- [67] D.L. Ergun, W.W. Pepler, J.T. Dobbins III, F.E. Zink, D.G. Kruger, F. Kelcz, F.J. de Bruijn, E.W. Bellers, Y. Wang, R.J. Althof, M.G.J. Wind, “Dual-energy computer radiography”, in *Proceedings SPIE*, vol. 2167, pp. 663–671, 1994. [20](#)
- [68] O.D. Faugeras & W.K. Pratt, “Decorrelation methods of texture feature extraction”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, no. 4, pp. 323–332, 1980. [76](#)
- [69] L.M.J. Florack, *Image Structure*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1997. [77](#), [79](#)
- [70] L.M.J. Florack, B.M. ter Haar Romeny, J.J. Koenderink, M.A. Viergever, “The Gaussian scale-space paradigm and the multiscale local jet”, *International Journal of Computer Vision*, vol. 18, no. 1, pp. 61–75, 1996. [61](#), [79](#), [102](#)
- [71] L.M.J. Florack, B.M. ter Haar Romeny, J.J. Koenderink, M.A. Viergever, “Cartesian differential invariants in scale-space”, *Journal of Mathematical Imaging and Vision*, vol. 3, no. 4, pp. 327–348, November 1993. [61](#), [79](#)
- [72] C.E. Floyd, E.F. Patz, J.Y. Lo, N.F. Vittitoe, L.E. Stambaugh, “Diffuse nodular lung disease on chest radiographs: a pilot study of characterization by fractal dimension”, *American Journal of Roentgenology*, vol. 167, pp. 1185–1187, 1996. [32](#)
- [73] W.T. Freeman & E.H. Adelson, “The design and use of steerable filters”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891–906, 1991. [45](#), [79](#)
- [74] J. Frija, E. de Kerviler, A-M. Zagdanski, C. Feger, P. Attal, M. Laval-Jeantet, “Radiologie numérique du thorax”, *Journal de Radiologie*, vol. 78, pp. 193–207, 1997. [20](#), [39](#)
- [75] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed., Academic Press, 1990. [85](#)

- [76] J. Gårding & T. Lindeberg, “Direct Computation of Shape Cues Based on Scale-adapted Spatial Derivative Operators”, *International Journal of Computer Vision*, vol. 17, no. 2, pp. 163–192, 1996. [71](#), [82](#), [92](#)
- [77] L.H. Garland, “On the scientific evaluation of diagnostic procedures”, *Radiology*, vol. 52, no. 3, pp. 309–328, 1949. [15](#), [36](#)
- [78] R.J. Geluk, “Digital equalization radiography”, in *Proceedings SPIE*, vol. 3659, pp. 471–477, 1999. [21](#)
- [79] M.L. Giger, K. Doi, H. MacMahon, “Image feature analysis and computer-aided diagnosis in digital radiography: automated detection of nodules in peripheral lung fields”, *Medical Physics*, vol. 15, no. 2, pp. 158–166, 1988. [29](#), [30](#), [30](#), [30](#), [30](#), [30](#), [31](#)
- [80] M.L. Giger, K. Doi, H. MacMahon, C.E. Metz, F-F. Yin, “Pulmonary nodules: computer-aided detection in digital chest images”, *Radiographics*, vol. 10, pp. 41–51, 1990. [31](#), [33](#)
- [81] B. van Ginneken, J.J. Koenderink, K.J. Dana, “Texture histograms as a function of illumination and viewing direction”, *International Journal of Computer Vision*, vol. 31, no. 2/3, pp. 169–184, 1999. [70](#), [70](#)
- [82] B. van Ginneken & B.M. ter Haar Romeny, “Automatic segmentation of lung fields in chest radiographs”, in *MICCAI'99*, no. 1679 in Lecture Notes in Computer Science, Springer, Berlin, pp. 184–191, 1999. [22](#), [23](#), [24](#), [119](#)
- [83] B. van Ginneken & B.M. ter Haar Romeny, “Automatic delineation of ribs in chest radiographs”, in *Proceedings SPIE*, vol. 3979, pp. 825–836, 2000. [25](#), [26](#)
- [84] B. van Ginneken & B.M. ter Haar Romeny, “Automatic segmentation of lung fields in chest radiographs”, *Medical Physics*, vol. 27, no. 10, pp. 2445–2455, 2000. [22](#), [23](#), [24](#), [119](#), [137](#)
- [85] J. Glassroth, “Diagnosis of tuberculosis”, in *Tuberculosis: a comprehensive international approach*, Marcel Dekker, Inc., pp. 149–165, 1999. [15](#)
- [86] C. Goodall, “Procrustes methods in the statistical analysis of shapes”, *Journal of the Royal Statistical Society B*, vol. 53, no. 2, pp. 285–339, 1991. [97](#)
- [87] L. van Gool, P. Dewaele, A. Oosterlinck, “Survey: Texture analysis anno 1983”, *Computer Vision, Graphics, and Image Processing*, vol. 29, pp. 336–357, 1985. [76](#), [134](#)
- [88] U. Grenander, Y. Chow, D. Keenan, *HANDS. A pattern theoretical study of biological shapes*, Springer Verlag, New York, 1991. [96](#)
- [89] L.D. Griffin, “Scale-imprecision space”, *Image and Vision Computing*, vol. 15, pp. 369–398, 1997. [61](#), [64](#), [77](#)
- [90] D.M. Guantt & G.T. Barnes, “X-ray tube potential, filtration, and detector considerations in dual-energy chest radiography”, *Medical Physics*, vol. 21, no. 2, pp. 203–218, 1994. [20](#)
- [91] F. Guichard & J.M. Morel, “Partial differential equations and image iterative filtering”, in *State of the Art in Numerical Analysis*, Oxford University Press, 1997. [64](#)
- [92] J.W. Gurney & S.J. Swensen, “Solitary pulmonary nodules: determining the likelihood of malignancy with neural network analysis”, *Radiology*, vol. 196, pp. 823–829, 1995. [33](#)
- [93] B.M. ter Haar Romeny (ed.), *Geometry-driven diffusion in computer vision*, Kluwer Academic Publishers, Dordrecht, 1994. [61](#)
- [94] B.M. ter Haar Romeny, L.M.J. Florack, A.H. Salden, M.A. Viergever, “Higher order differential structure of images”, *Image and Vision Computing*, vol. 12, no. 6, pp. 317–325, 1994. [61](#)
- [95] B.M. ter Haar Romeny, B. Titulaer, S. Kalitzin, G. Scheffer, F. Broekmans, E. te Velde, J.J. Staal, “Computer assisted human follicle analysis for fertility prospects with 3D ultrasound”, in *IPMI '99*, vol. 1613 of *Lecture Notes in Computer Science*, Springer-Verlag, Heidelberg, pp. 56–69, 1999. [96](#)
- [96] D.L. Hall, G.S. Lodwick, R.P. Kruger, S.J. Dwyer III, “Computer diagnosis of heart disease”, *Radiological Clinics of North America*, vol. 9, no. 3, pp. 533–541, 1971. [26](#), [28](#)

- [97] E.L. Hall, W.O. Crawford Jr., F.E. Roberts, "Computer classification of pneumoconiosis from radiographs of coal workers", *IEEE Transactions on Biomedical Engineering*, vol. BME-22, no. 6, pp. 518–527, 1975. [24](#), [34](#), [134](#)
- [98] E.L. Hall, R.P. Kruger, S.J. Dwayer III, D.L. Hall, R.W. McLaren, G.S. Lodwick, "A survey of pre-processing and feature extraction techniques for radiographic images", *IEEE Transactions on Computers*, vol. 20, no. 9, pp. 1032–1044, 1971. [21](#)
- [99] E.J. Hall, R.P. Kruger, A.F. Turner, "An optical-digital system for automatic processing of chest X-rays", *Optical Engineering*, vol. 13, no. 3, pp. 250–257, 1974. [34](#)
- [100] R.M. Haralick, "Statistical and structural approaches to texture", *Proc. of the IEEE*, vol. 67, no. 5, pp. 786–804, 1979. [34](#), [76](#), [82](#), [134](#)
- [101] R.M. Haralick, K. Shanmugam, I. Dinstein, "Textural Features for Image Classification", *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, pp. 610–621, 1973. [31](#), [34](#), [76](#), [79](#), [80](#), [134](#)
- [102] C.A. Harlow & S.A. Eisenbeis, "The analysis of radiographic images", *IEEE Transactions on Computers*, vol. 22, no. 7, pp. 678–689, 1973. [23](#)
- [103] A. Hasegawa, S.C. Lo, M.T. Freedman, S.K. Mun, "Convolution neural network based detection of lung structures", in *Proceedings SPIE*, vol. 2167, pp. 654–662, 1994. [23](#), [39](#)
- [104] D.C. He & L. Wang, "Unsupervised textural classification of images using the texture spectrum", *Pattern Recognition*, vol. 25, no. 3, pp. 247–255, 1992. [76](#)
- [105] A. Hill, T.F. Cootes, C.J. Taylor, "Active shape models and the shape approximation problem", *Image and Vision Computing*, vol. 14, no. 8, pp. 601–607, 1996. [101](#), [115](#)
- [106] Kundel H.L., C.F. Nodine, E.A. Krupinsky, "Searching for lung nodules: visual dwell indicates locations of false-positive and false-negative decisions", *Investigate Radiology*, vol. 24, pp. 424–427, 1989. [20](#)
- [107] J-T. Ho & R.A. Kruger, "Comparison of dual and single exposure techniques in dual-energy chest radiography", *Medical Physics*, vol. 16, no. 2, pp. 202–208, 1989. [20](#)
- [108] R.A. Hummel, "Image enhancement by histogram transformation", *Computer Graphics and Image Processing*, vol. 6, pp. 184–195, 1977. [68](#)
- [109] M. Ibrahim Sezan, A. Murat Tekalp, R. Schaetzing, "Automatic anatomically selective image enhancement in digital chest radiography", *IEEE Transactions on Medical Imaging*, vol. 8, no. 2, pp. 154–161, 1989. [21](#)
- [110] T. Ishida, K. Ashizawa, R. Engelmann, S. Katsuragawa, H. MacMahon, K. Doi, "Application of temporal subtraction for detection of interval changes on chest radiographs: improvement of subtraction images using automated initial image matching", *Journal of Digital Imaging*, vol. 12, no. 2, pp. 77–86, 1999. [22](#)
- [111] T. Ishida, S. Katsuragawa, K. Ashizawa, H. MacMahon, K. Doi, "Application of artificial neural networks for quantitative analysis of image data in chest radiographs for detection of interstitial lung disease", *Journal of Digital Imaging*, vol. 11, no. 4, pp. 182–192, 1998. [35](#), [35](#), [134](#), [151](#)
- [112] T. Ishida, S. Katsuragawa, T. Kobeyashi, H. MacMahon, K. Doi, "Computerized analysis of interstitial disease in chest radiographs: improvement of geometric-pattern feature analysis", *Medical Physics*, vol. 24, no. 6, pp. 915–924, 1997. [35](#)
- [113] T. Ishida, S. Katsuragawa, K. Nakamura, H. MacMahon, K. Doi, "Iterative image warping technique for temporal subtraction of sequential chest radiographs to detect interval change", *Medical Physics*, vol. 26, pp. 1320–1329, 1999. [22](#)
- [114] J.R. Jagoe, "Gradient pattern coding – An application to the measurement of pneumoconiosis in chest X-rays", *Computers and Biomedical Research*, vol. 12, pp. 1–15, 1979. [34](#), [134](#)
- [115] J.R. Jagoe & K.A. Paton, "Reading chest radiographs for pneumoconiosis by computer", *British Journal of Industrial Medicine*, vol. 32, pp. 267–272, 1975. [24](#), [34](#), [134](#)
- [116] J.R. Jagoe & K.A. Paton, "Measurement of pneumoconiosis by computer", *IEEE Transactions on Computers*, vol. 25, pp. 95–97, 1976. [34](#), [134](#)

- [117] A.K. Jain, *Fundamentals of digital image processing*, Prentice Hall, Englewood Cliffs, NJ, 1989. 33
- [118] A.K. Jain, R.P.W. Duin, J. Mao, “Statistical pattern recognition: a review”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000. 151, 159
- [119] A.K. Jain, Y. Zhong, M.P. Dubuisson-Jolly, “Deformable template models: a review”, *Signal Processing*, vol. 71, no. 2, pp. 109–129, 1996. 76, 96
- [120] A.K. Jain, Y. Zhong, S. Lakshaman, “Object matching using deformable templates”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 3, pp. 267–278, 1996. 96
- [121] I.T. Joliffe, *Principal Components Analysis*, Springer Verlag, New York, 1986. 120
- [122] B. Julesz, “Textons, the elements of texture perception and their interactions”, *Nature*, vol. 290, pp. 91–97, 1981. 82
- [123] B. Julesz, “A theory of preattentive texture discrimination based on first-order statistics of textons”, *Biological Cybernetics*, vol. 41, pp. 131–138, 1981. 82
- [124] A. Kano, K. Doi, H. MacMahon, D.D. Hassell, M.L. Giger, “Digital image subtraction of temporally sequential chest images for detection of interval change”, *Medical Physics*, vol. 21, no. 3, pp. 453–461, 1994. 22, 22, 34
- [125] M. Kass, A. Witkin, D. Terzopoulos, “Snakes: active contour models”, *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988. 96
- [126] S. Katsuragawa, K. Doi, H. MacMahon, “Image feature analysis and computer-aided diagnosis in digital radiography: detection and characterization of interstitial lung disease in digital chest radiographs”, *Medical Physics*, vol. 15, no. 3, pp. 311–319, 1988. 34, 134, 134
- [127] S. Katsuragawa, K. Doi, H. MacMahon, “Image feature analysis and computer-aided diagnosis in digital radiography: classification of normal and abnormal lungs with interstitial disease in chest images”, *Medical Physics*, vol. 16, no. 1, pp. 38–44, 1989. 34, 134, 134
- [128] S. Katsuragawa, K. Doi, H. MacMahon, L. Monnier-Cholley, T. Ishida, T. Kobayashi, “Classification of normal and abnormal lungs with interstitial diseases by rule-based method and artificial neural networks”, *Journal of Digital Imaging*, vol. 10, no. 3, pp. 108–114, 1997. 35, 134, 151
- [129] S. Katsuragawa, K. Doi, H. MacMahon, L. Monnier-Cholley, J. Morishita, T. Ishida, “Quantitative analysis of geometric pattern features of interstitial infiltrates in digital chest radiographs: preliminary results”, *Journal of Digital Imaging*, vol. 9, no. 3, pp. 137–144, 1996. 35, 35, 134
- [130] S. Katsuragawa, K. Doi, H. MacMahon, N. Nakamori, Y. Sasaki, J.J. Fennessy, “Quantitative computer-aided analysis of lung texture in chest radiographs”, *Radiographics*, vol. 10, no. 2, pp. 257–269, 1990. 35
- [131] S. Katsuragawa, H. Tagashira, Q. Li, H. MacMahon, K. Doi, “Comparison of the quality of temporal subtraction images obtained with manual and automated methods of digital chest radiography”, *Journal of Digital Imaging*, vol. 12, no. 4, pp. 166–172, 1999. 22
- [132] A. Kelemen, G. Székely, G. Gerig, “Elastic model-based segmentation of 3-D neuroradiological data sets”, *IEEE Transactions on Medical Imaging*, vol. 18, no. 10, pp. 828–839, 1999. 96
- [133] S. Kido, J. Ikezoe, H. Naito, M. Masuike, S. Tamura, T. Kozuka, “An image analyzing system for interstitial lung abnormalities in chest radiography”, *Radiographics*, vol. 15, pp. 1457–1464, 1995. 35, 134
- [134] S. Kido, J. Ikezoe, H. Naito, S. Tamura, S. Machi, “Fractal analysis of interstitial lung abnormalities in chest radiography”, *Radiographics*, vol. 15, pp. 1457–1464, 1995. 35, 134
- [135] T. Kobayashi, X-W. Xu, H. MacMahon, C. Metz, K. Doi, “Effect of a computer-aided diagnosis scheme on radiologists’ performance in detection of lung nodules on radiographs”, *Radiology*, vol. 199, pp. 843–848, 1996. 33

- [136] J.J. Koenderink, “The structure of images”, *Biological Cybernetics*, vol. 50, pp. 363–370, 1984. [61](#), [62](#), [77](#)
- [137] J.J. Koenderink & A.J. van Doorn, “Representation of local geometry in the visual system”, *Biological Cybernetics*, vol. 55, pp. 367–375, 1987. [61](#), [79](#), [102](#)
- [138] J.J. Koenderink & A.J. Van Doorn, “Operational significance of receptive field assemblies”, *Biological Cybernetics*, vol. 58, pp. 163–171, 1988. [61](#)
- [139] J.J. Koenderink & A.J. van Doorn, “Receptive field families”, *Biological Cybernetics*, vol. 63, pp. 291–298, 1990. [61](#)
- [140] J.J. Koenderink & A.J. van Doorn, “Generic neighborhood operators”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 6, pp. 597–605, 1992. [61](#)
- [141] J.J. Koenderink & A.J. van Doorn, “The structure of locally orderless images”, *International Journal of Computer Vision*, vol. 31, no. 2/3, pp. 159–168, 1999. [61](#), [61](#), [62](#), [75](#), [77](#), [102](#), [156](#)
- [142] R.P. Kruger, W.B. Thompson, A.F. Turner, “Computer diagnosis of pneumoconiosis”, *IEEE Transactions on Systems, Man and Cybernetics*, vol. 4, p. 40, 1974. [34](#), [134](#)
- [143] R.P. Kruger, J.R. Townes, D.L. Hall, S.J. Dwyer III, G.S. Lodwick, “Automated radiographic diagnosis via feature extraction and classification of cardiac size and shape descriptors”, *IEEE Biomedical Transactions*, vol. BME-19, no. 3, pp. 174–186, 1972. [24](#), [26](#), [28](#), [28](#), [29](#)
- [144] E.A. Krupinsky, C.F. Nodine, H.L. Kundel, “A perceptually based method for enhancing pulmonary nodule recognition”, *Investigative Radiology*, vol. 28, no. 4, pp. 289–294, 1993. [20](#)
- [145] H.L. Kundel, G. Revesz, L. Toto, “Contrast gradient and the detection of lung nodules”, *Investigative Radiology*, vol. 14, pp. 18–22, 1979. [33](#)
- [146] A. Laine & J. Fan, “Texture classification by wavelet packet signatures”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1186–1191, 1993. [76](#)
- [147] W.A. Lampeter, “ANDS-V1 computer detection of lung nodules”, in *Proceedings SPIE*, vol. 555, pp. 253–261, 1985. [29](#), [29](#), [31](#)
- [148] W.A. Lampeter & J.C. Wandtke, “Computerized search of chest radiographs for nodules”, *Investigative Radiology*, vol. 21, pp. 384–390, 1986. [29](#), [30](#), [31](#)
- [149] S. Lange, *Radiology of Chest Diseases*, Georg Thieme Verlag, Stuttgart, 1990. [21](#)
- [150] K.D. Laws, *Texture image segmentation*, Ph.D. thesis, University of Southern California, 1980. [31](#), [76](#), [87](#), [134](#)
- [151] R.S. Ledley, H.K. Huang, L.S. Rotola, “A texture analysis method in classification of coal workers’ pneumoconiosis”, *Computers in Biology and Medicine*, vol. 5, pp. 53–67, 1975. [34](#)
- [152] C.C. Li & A.M. Savol, “Computer analysis of small opacities in coal miners’ chest X-rays”, in *Proceedings of the IEEE International Conference on Cybernetics and Society*, pp. 422–428, 1977. [34](#)
- [153] C.C. Li, A.M. Savol, C.P. Fong, “An improved algorithm for detection of small rounded pneumoconiosis opacities in chest X-rays”, in *??*, pp. 390–395, 1978. [34](#)
- [154] Q. Li, S. Katsuragawa, K. Doi, “Improved contralateral subtraction images by use of elastic matching technique”, *Medical Physics*, vol. 27, no. 8, pp. 1934–1942, 2000. [22](#)
- [155] Q. Li, S. Katsuragawa, T. Ishida, H. Yoshida, S. Tsukuda, H. MacMahon, K. Doi, “Contralateral subtraction: a novel technique for detection of asymmetric abnormalities on digital chest radiographs”, *Medical Physics*, vol. 27, no. 1, pp. 47–55, 2000. [22](#), [134](#), [150](#)
- [156] B. Likar, J.B.A. Maintz, M.A. Viergever, F. Pernuš, “Retrospective shading correction based on entropy minimisation”, *Journal of Microscopy*, vol. 197, pp. 285–295, 2000. [70](#)
- [157] B. Likar, M.A. Viergever, F. Pernuš, “Retrospective correction of MR intensity inhomogeneity by information minimization”, in *MICCAI*, no. 1935 in Lecture Notes in Computer Science, Springer, Berlin, pp. 375–384, 2000. [70](#)
- [158] J-S. Lin, A. Hasegawa, M.T. Freedman, S.K. Mun, “Differentiation between nodules and end-on vessels using a convolution neural network architecture”, *Journal of Digital Imaging*, vol. 8, no. 3, pp. 132–141, 1995. [32](#)

- [159] J-S. Lin, S.C. Lo, A. Hasegawa, M.T. Freedman, S.K. Mun, "Reduction of false positives in lung nodule detection using a two-level neural classification", *IEEE Transactions on Medical Imaging*, vol. 15, no. 2, pp. 206–217, 1996. [32](#), [32](#), [33](#)
- [160] T. Lindeberg, *Scale-Space Theory in Computer Vision*, Kluwer Academic Publishers, Dordrecht, the Netherlands, 1994. [62](#), [77](#)
- [161] T. Lindeberg, "Edge detection and ridge detection with automatic scale selection", *International Journal of Computer Vision*, vol. 30, no. 2, pp. 117–154, 1998. [62](#)
- [162] S. Liu & M.E. Jernigan, "Texture analysis and discrimination in additive noise", *Computer Vision, Graphics, and Image Processing*, vol. 49, pp. 52–67, 1990. [76](#)
- [163] S-C.B. Lo, M.T. Freedman, J-S. Lin, S.K. Mun, "Automatic lung nodule detection using profile matching and back-propagation neural network techniques", *Journal of Digital Imaging*, vol. 6, no. 1, pp. 48–54, 1993. [29](#), [30](#)
- [164] S-C.B. Lo, S-L.A. Lou, J-S. Lin, M.T. Freedman, M.V. Chien, S.K. Mun, "Artificial convolution neural network techniques and applications for lung nodule detection", *IEEE Transactions on Medical Imaging*, vol. 14, no. 4, pp. 711–718, 1995. [31](#), [33](#)
- [165] G.S. Lodwick, T.E. Keats, J.P. Dorst, "The coding of Roentgen images for computer analysis as applied to lung cancer", *Radiology*, vol. 81, no. 2, pp. 185–200, 1963. [20](#)
- [166] L-N. Loo, K. Doi, C.E. Metz, "Investigation of basic imaging properties in digital radiography. 4. Effect of unsharp masking on the detectability of simple patterns", *Medical Physics*, vol. 12, no. 2, pp. 209–214, 1985. [20](#)
- [167] H. MacMahon, R. Engelmann, F.M. Behlen, K.R. Hoffmann, T. Ishida, C. Roe, C.E. Metz, K. Doi, "Computer-aided diagnosis of pulmonary nodules: results of a large-scale observer test", *Radiology*, vol. 213, pp. 723–726, 1999. [33](#)
- [168] H. MacMahon, K.J.M. Liu, S.M. Montner, K. Doi, "The nature and subtlety of abnormal findings in chest radiographs", *Medical Physics*, vol. 18, no. 2, pp. 206–210, 1991. [27](#), [28](#), [35](#), [37](#)
- [169] H. MacMahon, C.J. Vyborny, C.E. Metz, K. Doi, V. Sabeti, S.L. Solomon, "Digital radiography of subtle pulmonary abnormalities: an ROC study of the effect of pixel size on observer performance", *Radiology*, vol. 158, pp. 21–26, 1986. [20](#)
- [170] H. MacMahon, C. Vyborny, V. Sabeti, C.E. Metz, K. Doi, "The effect of digital unsharp masking on the detectability of interstitial infiltrates and pneumothoraces", in *Proceedings SPIE*, vol. 555, pp. 246–252, 1985. [20](#)
- [171] D. Maher & M.C. Raviglione, "The global epidemic of tuberculosis: a World Health Organization perspective", in *Tuberculosis and nontuberculous Mycobacterial infections*, 4 ed., W.B. Saunders Company, pp. 104–115, 1999. [14](#)
- [172] J. Malik & P. Perona, "Preattentive Texture Discrimination With Early Vision Mechanisms", *Journal of the Optical Society of America-A*, vol. 7, no. 5, pp. 923–932, 1990. [76](#), [134](#)
- [173] F. Mao, W. Qian, J. Gaviria, L.P. Clarke, "Fragmentary window filtering for multiscale lung nodule detection", *Academic Radiology*, vol. 5, no. 4, pp. 306–311, 1998. [29](#), [30](#)
- [174] T. Matsumoto, H. Yoshimura, K. Doi, M.L. Giger, A. Kano, H. MacMahon, M. Abe, S.M. Montner, "Image feature analysis of false-positive diagnoses produced by automated detection of lung nodules", *Investigative Radiology*, vol. 27, no. 8, pp. 587–579, 1992. [29](#), [31](#), [31](#)
- [175] T. Matsumoto, H. Yoshimura, M.L. Giger, K. Doi, A. Kano, H. MacMahon, M. Abe, S. Montner, "Potential usefulness of computerized nodule detection in screening programs for lung cancer", *Investigative Radiology*, vol. 27, pp. 471–475, 1992. [33](#)
- [176] H.P. McAdams, G.A. Johnson, S.A. Suddarth, C.E. Ravin, "Histograms-directed processing of digital chest images", *Investigative Radiology*, vol. 21, pp. 253–259, 1986. [21](#)
- [177] T. McInerney & D. Terzopoulos, "Deformable models in medical image analysis: a survey", *Medical Image Analysis*, vol. 1, no. 2, pp. 91–108, 1996. [96](#)

- [178] M.F. McNitt-Gray, H.K. Huang, J.W. Sayre, "Feature selection in the pattern classification problem of digital chest radiograph segmentation", *IEEE Transactions on Medical Imaging*, vol. 14, no. 3, pp. 537–547, 1995. [23](#), [23](#), [24](#), [39](#), [53](#), [53](#), [55](#), [137](#)
- [179] M.F. McNitt-Gray, J.W. Sayre, H.K. Huang, M. Razavi, "A pattern classification approach to segmentation of chest radiographs", in *Proceedings SPIE*, vol. 1898, pp. 160–170, 1993. [23](#), [24](#), [39](#), [55](#)
- [180] M.F. McNitt-Gray, R.K. Taira, S.L. Eldredge, M. Razavi, "Brightness and contrast adjustments for different tissue densities in digital chest radiographs", in *Proceedings SPIE*, vol. 1445, pp. 468–478, 1991. [21](#)
- [181] C.E. Metz, "ROC methodology in radiologic imaging", *Investigative Radiology*, vol. 21, no. 9, pp. 720–733, 1986. [141](#)
- [182] P.H. Meyers, C.M. Nice Jr., H.C. Becker, W.J. Nettleton, J.W. Sweeney, G.R. Meckstroth, "Automated computer analysis of radiographic images", *Radiology*, vol. 83, pp. 1029–1034, 1964. [19](#), [26](#), [28](#)
- [183] L. Monnier-Cholley, H. MacMahon, S. Katsuragawa, J. Morishita, T. Ishida, K. Doi, "Computerized analysis of interstitial infiltrates on chest radiographs: a new scheme based on geometric pattern features and Fourier analysis", *Academic Radiology*, vol. 2, pp. 455–462, 1995. [35](#)
- [184] L. Monnier-Cholley, H. MacMahon, S. Katsuragawa, J. Morishita, T. Ishida, K. Doi, "Computer-aided diagnosis for detection of interstitial opacities on chest radiographs", *American Journal of Roentgenology*, vol. 171, pp. 1651–1656, 1998. [35](#), [134](#)
- [185] J. Morishita, K. Doi, S. Katsuragawa, L. Monnier-Cholley, H. MacMahon, "Computer-aided diagnosis for interstitial infiltrates in chest radiographs: optical-density dependence of texture measures", *Medical Physics*, vol. 22, no. 9, pp. 1515–1522, 1995. [35](#), [134](#)
- [186] J.R. Muhm, W.E. Miller, R.S. Fontana, D.R. Sanderson, M.A. Uhlenhopp, "Lung cancer detected during a screening program using four-month chest radiographs", *Radiology*, vol. 148, pp. 609–615, 1983. [29](#)
- [187] R. Mukundan & K.R. Ramakrishnan, *Moment Functions in Image Analysis*, World Scientific, Singapore, 1998. [80](#)
- [188] N. Nakamori, K. Doi, H. MacMahon, Y. Sasaki, S.M. Montner, "Effect of heart-size parameters computed from digital chest radiographs on detection of cardiomegaly: potential usefulness for computer-aided diagnosis", *Investigative Radiology*, vol. 26, no. 6, pp. 546–550, 1991. [26](#), [29](#)
- [189] N. Nakamori, K. Doi, V. Sabeti, H. MacMahon, "Image feature analysis and computer-aided diagnosis in digital radiography: automated analysis of sizes of heart and lung in chest images", *Medical Physics*, vol. 17, no. 3, pp. 342–350, 1990. [26](#), [28](#)
- [190] K. Nakamura, H. Yoshida, R. Engelmann, H. MacMahon, S. Katsuragawa, T. Ishida, K. Ahizawa, K. Doi, "Computerized analysis of the likelihood of malignancy in solitary pulmonary nodules with use of artificial neural networks", *Radiology*, vol. 214, pp. 823–830, 2000. [32](#), [33](#)
- [191] P.P. Ohanian & R.C. Dubes, "Performance evaluation for four classes of textural features", *Pattern Recognition*, vol. 25, no. 8, pp. 819–833, 1992. [76](#)
- [192] E. Oja & K. Valkealahti, "Co-occurrence map: quantizing multi-dimensional texture histograms", *Pattern Recognition Letters*, vol. 17, pp. 723–730, 1996. [77](#), [80](#)
- [193] World Health Organization, *The World Health Report 1998, Life in the 21st century - A vision for all*, 1998. [14](#), [133](#)
- [194] J.L. Paul, M.D. Levine, R.G. Fraser, C.A. Laszlo, "The measurement of total lung capacity based on a computer analysis of anterior and lateral radiographic chest images", *IEEE Transactions on Biomedical Engineering*, vol. BME-21, no. 6, pp. 444–451, 1974. [24](#), [29](#)
- [195] P. Perona & J. Malik, "Scale-space and edge detection using anisotropic diffusion", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990. [61](#)

- [196] Y.S. Peter Chiou, J-S.J. Lin, Y.M. Fleming Lure, P.A. Ligomenides, M.T. Freedman, S. Fritz, "Shape feature analysis using artificial neural networks for improvements of hybrid lung nodule detection (HLND) system", in *Proceedings SPIE*, vol. 1898, pp. 609–617, 1993. 31
- [197] M. Pietikäinen, A. Rosenfeld, L.S. Davis, "Experiments with texture classification using averages of local pattern matches", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 13, no. 3, pp. 421–426, 1983. 76
- [198] E. Pietka, "Lung segmentation in digital chest radiographs", *Journal of Digital Imaging*, vol. 2, pp. 79–84, 1994. 22, 23, 137
- [199] S.M. Pizer, E.P. Amburn, J.D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B.M. ter Haar Romeny, J. Zimmerman, K. Zuiderveld, "Adaptive histogram equalization and its variations", *Computer Vision, Graphics, and Image Processing*, vol. 39, pp. 355–368, 1987. 69
- [200] S.M. Pizer, D.S. Fritsch, P.A. Yushkevich, V.E. Johnson, E.L. Chaney, "Segmentation, registration, and measurement of shape variation via image object shape", *IEEE Transactions on Medical Imaging*, vol. 18, no. 10, pp. 851–865, 1999. 97
- [201] G.F. Powell, K. Doi, S. Katsuragawa, "Localization of inter-rib spaces for lung texture analysis and computer-aided diagnosis in digital chest images", *Medical Physics*, vol. 15, no. 4, pp. 581–587, 1988. 24, 25, 25, 118
- [202] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipes in C; the Art of Scientific Computing*, 2 ed., Cambridge University Press, Cambridge, UK, 1992. 49, 125
- [203] P. Pudil, J. Novovicova, J. Kittler, "Floating search methods in feature selection", *Pattern Recognition Letters*, vol. 15, no. 11, pp. 1119–1125, 1994. 85, 103
- [204] R.T. Reed & J.M.H. du Buf, "A review of recent texture segmentation and feature extraction techniques", *Computer Vision, Graphics, and Image Processing*, vol. 57, no. 3, pp. 359–372, 1993. 76, 134
- [205] G. Revesz & H.L. Kundel, "Feasibility of classifying disseminated pulmonary diseases based on their Fourier spectra", *Investigative Radiology*, vol. 8, pp. 345–349, 1973. 34, 134
- [206] F.X. Roellinger, A.E. Kahveci, J.K. Chang, C.A. Harlow, S.J. Dwyer III, G.S. Lodwick, "Computer analysis of radiographic images", *Computer Graphics and Image Processing*, vol. 2, pp. 232–251, 1973. 28
- [207] R. Ronfard, "Region-based strategies for active contour models", *International Journal of Computer Vision*, vol. 13, no. 2, pp. 229–251, 1994. 96
- [208] W.C. Röntgen, "Ueber eine neue Art von Strahlen", *Sitzungsberichte der Physikalisch-Medicinisch Gesellschaft zu Würzburg*, pp. 132–141, 1895. 15, 19
- [209] D. Ruprecht & H. Müller, "Image Warping with Scattered Data Interpolation", *IEEE Computer Graphics and Applications*, vol. 15, no. 2, pp. 37–46, 1995. 139
- [210] E. Samei, *The performance of digital X-ray imaging systems in detection of subtle lung nodules*, Ph.D. thesis, Univeristy of Michigan, Ann Arbor, May 1997. 20
- [211] S. Sanada, K. Doi, H. MacMahon, "Image feature analysis and computer-aided diagnosis in digital radiography: automated delineation of posterior ribs in chest images", *Medical Physics*, vol. 18, no. 5, pp. 964–971, 1991. 25, 118
- [212] S Sanada, K. Doi, H. MacMahon, "Image feature analysis and computer-aided diagnosis in digital radiography: automated detection of pneumothorax in chest images", *Medical Physics*, vol. 19, no. 5, pp. 1153–1160, 1992. 35
- [213] P.V. Sankar & J. Sklansky, "A Gestalt guided heuristic boundary follower for X-ray images of lung nodules", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 4, no. 3, pp. 326–331, 1982. 31
- [214] C. Schaefer-Prokop & M. Prokop, "Digital radiography of the chest: comparison of the selenium detector with other imaging systems", *MedicaMundi*, vol. 41, no. 1, pp. 2–11, 1997. 20
- [215] S. Sclaroff & J. Isodor, "Active blobs", in *Proceedings of the IEEE ICCV*, pp. 1146–1153, 1998. 97

- [216] S. Sclaroff & A.P. Pentland, “Modal matching for correspondence and recognition”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 7, pp. 545–561, 1995. [97](#)
- [217] N. Sezaki & K. Ukena, “Automatic computation of the cardiothoracic ratio with application to mass screening”, *IEEE Transactions on Biomedical Engineering*, vol. BME-20, no. 4, pp. 248–253, 1973. [26](#), [28](#)
- [218] R.H. Sherrier & G.A. Johnson, “Regionally adaptive histogram equalization of the chest”, *IEEE Transactions on Medical Imaging*, vol. 6, no. 1, pp. 1–7, 1987. [21](#)
- [219] R.H. Sherrier, G.A. Johnson, S.A. Suddarth, C. Chiles, C. Hulka, C.E. Ravin, “Digital synthesis of lung nodules”, *Investigative Radiology*, vol. 20, no. 9, pp. 933–937, 1985. [33](#)
- [220] J. Shi & J. Malik, “Normalized Cuts and Image Segmentation”, *PAMI*, vol. 22, no. 8, pp. 888–905, 2000. [71](#)
- [221] J. Sklansky & D. Ballard, “Tumor detection in radiographs”, *Computers and Biomedical Research*, vol. 6, no. 4, pp. 299–321, 1973. [30](#), [36](#)
- [222] J. Sklansky & D. Petković, “Two-resolution detection of lung tumors in chest radiographs”, in *Multiresolution Image Processing and Analysis*, A. Rosenfeld (ed.), Springer, Berlin, pp. 365–378, 1984. [30](#), [31](#)
- [223] P. Somol, P. Pudil, J. Novovicova, P. Paclik, “Adaptive floating search methods in feature selection”, *Pattern Recognition Letters*, vol. 20, no. 11–13, pp. 1157–1163, 1999. [85](#), [103](#)
- [224] P. Sozou, T.F. Cootes, C.J. Taylor, E. DiMauro, “A non-linear generalisation of point distribution models using polynomial regression”, *Image and Vision Computing*, vol. 13, no. 5, pp. 451–457, 1995. [100](#)
- [225] L.F. Squire & R.A. Novelline, *Fundamentals of Radiology*, 4 ed., Harvard University Press, 1988. [21](#), [40](#), [117](#), [119](#)
- [226] H. Stark & D. Lee, “An optical-digital approach to the pattern recognition of coal workers’ pneumoconiosis”, *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-6, no. 11, pp. 788–793, 1976. [34](#)
- [227] R.N. Sutton & E.L. Hall, “Texture measures for automatic classification of pulmonary disease”, *IEEE Trans. Comput.*, vol. 21, pp. 667–676, 1972. [24](#), [34](#), [134](#)
- [228] H. Suzuki & N. Inaoka, “Development of a computer-aided detection system for lung cancer diagnosis”, in *Proceedings SPIE*, vol. 1652, pp. 567–571, 1992. [30](#), [31](#)
- [229] H. Suzuki, N. Inaoka, H. Takabatake, M. Mori, H. Natori, A. Suzuki, “An experimental system for detecting lung nodules by chest X-ray image processing”, in *Proceedings SPIE*, vol. 1450, pp. 99–107, 1991. [30](#), [31](#)
- [230] J. Suzuki, I. Furukawa, S. Ono, M. Kitamura, Y. Ando, “Contrast mapping and evaluation for electronic X-ray images on CRT display monitor”, *IEEE Transactions on Medical Imaging*, vol. 16, no. 6, pp. 772–784, 1997. [21](#)
- [231] A.W. Templeton, C. Jansen, J.L. Lehr, R. Hufft, “Solitary pulmonary lesions”, *Radiology*, vol. 89, no. 4, pp. 605–613, 1967. [20](#)
- [232] J. Toriwaki, J. Hasegawa, T. Fukumura, Y. Takagi, “Computer analysis of chest photofluorograms and its applications to automated screening”, *Automedica*, vol. 3, pp. 63–81, 1980. [23](#), [25](#), [26](#), [26](#)
- [233] J. Toriwaki, Y. Suenaga, T. Negoro, T. Fukumura, “Pattern recognition of chest X-ray images”, *Computer Graphics and Image Processing*, vol. 2, pp. 252–271, 1973. [23](#), [25](#), [26](#), [26](#), [118](#)
- [234] G.D. Tourassi, “Journey toward computer-aided diagnosis: role of image texture analysis”, *Radiology*, vol. 213, pp. 317–320, 1999. [134](#)
- [235] O. Tsujii, M.T. Freedman, S.K. Mun, “Automated segmentation of anatomic regions in chest radiographs using an adaptive-sized hybrid neural network”, *Medical Physics*, vol. 25, no. 6, pp. 998–1007, 1998. [23](#), [24](#), [39](#), [53](#), [57](#), [137](#)

- [236] R.J. Tully, R.W. Connors, C.A. Harlow, G.N. Larsen, S.J. Dwyer III, G.S. Lodwick, “Interactive analysis of pulmonary infiltration”, in *Proceedings of the 3rd International Joint Conference on Pattern Recognition*, pp. 238–242, 1976. 34
- [237] R.J. Tully, R.W. Connors, C.A. Harlow, G.S. Lodwick, “Towards computer analysis of pulmonary infiltration”, *Investigative Radiology*, vol. 13, pp. 298–305, 1978. 34, 134
- [238] A.M. Turing, “Can a machine think?”, in *Mind*, Oxford University Press, 1950. 11
- [239] A.F. Turner, R.P. Kruger, W.B. Thompson, “Automated computer screening of chest radiographs for pneumoconiosis”, *Investigative Radiology*, vol. 11, no. 4, pp. 258–266, 1976. 34
- [240] M.R. Turner, “Texture discrimination by Gabor functions”, *Biological Cybernetics*, vol. 55, pp. 71–82, 1986. 76
- [241] M. Unser, “Local linear transforms for texture measurements”, *Signal Processing*, vol. 11, pp. 61–79, 1986. 76, 87, 134
- [242] M. Unser, “Sum and difference histograms for texture classification”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 118–125, 1986. 76
- [243] M. Unser & M. Eden, “Multi-resolution feature extraction and selection for texture segmentation”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 717–728, 1989. 76, 134
- [244] K. Valkealahti & E. Oja, “Reduced multi-dimensional co-occurrence histograms in texture classification”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 90–94, 1998. 77, 80
- [245] J.F. Veenland, J.L. Grashuis, E.S. Gelsema, “Texture analysis in radiographs: the influence of modulation transfer function and noise on the discriminative ability of texture features”, *Medical Physics*, vol. 25, no. 6, pp. 922–936, 1998. 35
- [246] J.F. Veenland, J.L. Grashuis, F. van der Meer, A.L.D. Beckers, E.S. Gelsema, “Estimation of fractal dimension in radiographs”, *Medical Physics*, vol. 23, no. 4, pp. 585–594, 1996. 35
- [247] A.L. Vickers & J.W. Modestino, “A maximum likelihood approach to texture classification”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 1, no. 1, pp. 61–68, 1982. 76, 80
- [248] Vision & MIT Media Lab Modeling group, “VisTex database, <http://wwwwhite.media.mit.edu/vismod/imagery/VisionTexture/vistex.html>. 84
- [249] N.F. Vittitoe, J.A. Baker, C.E. Floyd Jr., “Fractal texture analysis in computer-aided diagnosis of solitary pulmonary nodules”, *Academic Radiology*, vol. 4, pp. 96–101, 1997. 32
- [250] N.F. Vittitoe, R. Vargas-Voracek, C.E. Floyd Jr., “Identification of lung regions in chest radiographs using Markov Random Field modeling”, *Medical Physics*, vol. 25, no. 6, pp. 976–985, 1998. 23, 24, 39, 53, 53, 53, 56, 137
- [251] N.F. Vittitoe, R. Vargas-Voracek, C.E. Floyd Jr., “Markov Random Field modeling in posteroanterior chest radiograph segmentation”, *Medical Physics*, vol. 26, no. 8, pp. 1670–1677, 1999. 23, 23, 24, 39, 137
- [252] F. Vogelsang, M. Kohnen, J. Mahlke, F. Weiler, M.W. Kilbinger, B. Wein, R.W. Günther, “Model based analysis of chest radiographs”, in *Proceedings SPIE*, vol. 3979, pp. 1040–1052, 2000. 22, 25, 26, 35
- [253] F. Vogelsang, F. Weiler, J. Dahmen, M.W. Kilbinger, B. Wein, R.W. Günther, “Detection and compensation of rib structures in chest radiographs for diagnose assistance”, in *Proceedings SPIE*, vol. 3338, pp. 774–785, 1998. 22, 25, 118, 134, 150
- [254] R.-D. von Müller, H. Hirche, M. Voß, B. Buddenbrock, V. John, P. Gocke, “ROC-Analyse zur Bildvernacharbeitung digitaler Thoraxaufnahmen”, *Fortschr. Röntgenstr.*, vol. 162, no. 2, pp. 163–169, 1995. 20
- [255] A.G. Weber, “The ISC-SIPI Image Database”, Tech. Rep., University of Southern California, 1997. <http://sipi.usc.edu>. 84

- [256] H. Wechsler, *Automatic detection of rib contours in chest radiographs*, Birkhauser Verlag, New York, 1977. 25, 118
- [257] H. Wechsler, “Image processing algorithms applied to rib boundary detection in chest radiographs”, *Computer Graphics and Image Processing*, vol. 7, pp. 375–390, 1978. 25, 26, 118
- [258] H. Wechsler & J. Sklansky, “Finding the rib cage in chest radiographs”, *Pattern Recognition*, vol. 9, pp. 21–30, 1977. 25
- [259] J. Wei, Y. Hagihara, H. Kobatake, “Detection of cancerous tumors on chest X-ray images – candidate detection filter and its evaluation”, in *Proceedings of the International Conference on Image Processing (ICIP’99, Kobe, Japan)*, IEEE Computer Society Press, Los Alamitos, CA, 1999. paper number 27AP4.2. 30
- [260] J. Weickert, “Multiscale Texture Enhancement”, in *Computer Analysis of Images and Patterns*, V. Hlaváč & R. Šára (eds.), vol. 970 of *Lecture Notes in Computer Science*, Springer, Berlin, pp. 230–237, 1995. 92
- [261] J. Weickert, S. Ishikawa, A. Imiya, “On the history of Gaussian scale-space axiomatics”, in *Gaussian Scale-Space Theory*, J. Sparring, M. Nielsen, L. Florack, P. Johansen (eds.), Kluwer, Dordrecht, pp. 45–59, 1997. 62
- [262] W.J. Weiser, “Sorting out interstitial lung disease”, *Applied Radiology*, pp. 58–65, April 1993. 21, 34
- [263] R. Weissleder, M.J. Rieumont, J. Wittenberg, *Primer of Diagnostic Imaging*, 2 ed., Mosby-Year Book, St. Louis, Missouri, 1997. 21
- [264] A. Whitney, “A direct method of non parametric measurement selection”, *IEEE Transactions on Computing*, vol. 20, pp. 1100–1103, 1971. 85, 103
- [265] L.S. Wilson, M.S. Brown, B.D. Doust, R.W. Gill, S.F. Brown, D. Nair, “Computer-aided diagnosis using anatomical models”, *Medical Imaging Technology*, vol. 6, no. 14, pp. 652–663, 1996. 23
- [266] Y.C. Wu, K. Doi, M.L. Giger, C.E. Metz, W. Zhang, “Reduction of false positives in computerized detection of lung nodules in chest radiographs using artificial neural networks, discriminant analysis, and a rule-based scheme”, *Journal of Digital Imaging*, vol. 7, no. 4, pp. 196–207, 1994. 31
- [267] Y.N. Wu, S.C. Zhu, X. Liu, “Equivalence of Julesz and Gibbs texture ensembles”, in *Proceedings of the ICCV*, pp. 1025–1032, 1999. 79
- [268] Y.N. Wu, S.C. Zhu, X. Liu, “Equivalence of Julesz and FRAME models”, *International Journal of Computer Vision*, vol. 38, no. 3, pp. 245–261, 2000. 77
- [269] X.W. Xu & K. Doi, “Image feature analysis for computer-aided diagnosis: accurate determination of ribcage boundary in chest radiographs”, *Medical Physics*, vol. 22, no. 5, pp. 617–626, 1995. 22, 22, 23, 39, 137
- [270] X.W. Xu & K. Doi, “Image feature analysis for computer-aided diagnosis: detection of right and left hemidiaphragm edges and delineation of lung field in chest radiographs”, *Medical Physics*, vol. 23, no. 9, pp. 1613–1624, 1996. 22, 22, 23, 39, 56, 137
- [271] X.W. Xu, S. Katsuragawa, K. Ashizawa, H. MacMahon, K. Doi, “Analysis of image features of histograms of edge gradient for false positive reduction in lung nodule detection in chest radiographs”, in *Proceedings SPIE*, vol. 3338, pp. 318–326, 1998. 29, 32
- [272] X.W. Xu, H. MacMahon, M.L. Giger, K. Doi, “Adaptive feature analysis of false positives for computerized detection of lung nodules in digital chest radiographs”, in *Proceedings SPIE*, vol. 3034, pp. 428–436, 1997. 32
- [273] J. Yerushalmy, “The statistical assessment of the variability in observer perception and description of Roentgenographic pulmonary shadows”, *Radiologic Clinics of North America*, vol. 7, no. 3, pp. 381–392, 1969. 15, 36

- [274] H. Yoshida, “Multiresolution non-rigid image registration method and its application to removal of normal anatomical structures in chest radiographs”, in *Proceedings of the International Conference on Image Processing (ICIP’99, Kobe, Japan)*, IEEE Computer Society Press, Los Alamitos, CA, 1999. Paper number 27AP4.11. [22](#)
- [275] H. Yoshida & K. Doi, “Computerized detection of pulmonary nodules in chest radiographs: reduction of false positives based on symmetry between left and right lungs”, in *Proceedings SPIE*, vol. 3979, pp. 97–102, 2000. [22](#), [32](#)
- [276] H. Yoshida, X.W. Xu, K. Doi, M.L. Giger, “Computer-aided diagnosis (CAD) scheme for detecting pulmonary nodules using wavelet transforms”, in *Proceedings SPIE*, vol. 2434, pp. 621–626, 1995. [30](#)
- [277] H. Yoshimura, M.L. Giger, K. Doi, H. MacMahon, S.M. Montner, “Computerized scheme for the detection of pulmonary nodules: a nonlinear filtering technique”, *Investigative Radiology*, vol. 27, pp. 124–127, 1992. [30](#)
- [278] Z. Yue, A. Goshtasby, L.V. Ackerman, “Automatic detection of rib borders in chest radiographs”, *IEEE Transactions on Medical Imaging*, vol. 14, no. 3, pp. 525–536, 1995. [24](#), [25](#), [25](#), [26](#), [118](#)
- [279] A.L. Yuille, P.W. Hallinan, D.S. Cohen, “Feature extraction from faces using deformable templates”, *International Journal of Computer Vision*, vol. 8, no. 2, pp. 99–112, 1992. [96](#)
- [280] S.C. Zhu, Y.N. Wu, D. Mumford, “Minimax entropy principle and its application to texture modeling”, *Neural Computation*, vol. 9, no. 8, pp. 1627–1660, 1997. [76](#), [77](#), [80](#)
- [281] S.W. Zucker & D. Terzopoulos, “Finding structure in co-occurrence matrices for texture analysis”, *Computer Graphics and Image Processing*, vol. 12, pp. 286–308, 1980. [76](#)

Samenvatting

Computers zijn onmisbaar geworden in onze maatschappij, maar op veel terreinen zijn mensen de snelste computers moeiteloos de baas. Het analyseren van beelden is zo'n terrein. Vandaar dat er veel onderzoek plaatsvindt naar beeldverwerking. Dit proefschrift is daar een voorbeeld van en gaat over computerondersteunde diagnose van Röntgenfoto's van de borstkas (thorax). De toepassing die centraal staat is tuberculosebestrijding. Thoraxfoto's vormen een belangrijk middel om gevallen van actieve tuberculose (TB) op te sporen en in minder ontwikkelde landen, waar TB jaarlijks drie miljoen slachtoffers eist, is vaak onvoldoende radiologische expertise aanwezig om grootschalig te kunnen screenen.

Twee rode draden lopen door het proefschrift: segmentatie en textuuranalyse. Segmentatie is nodig om te bepalen waar specifieke gedeelten van de longvelden en andere anatomische structuren zoals ribben zich in het beeld bevinden. Vervolgens kunnen dan de patronen (textuur) in elk gebied gekarakteriseerd worden. Na een inleiding in hoofdstuk 1 en een overzicht van de literatuur op het gebied van computeranalyse van thoraxfoto's in hoofdstuk 2, wordt in hoofdstuk 3 een methode beschreven om de longvelden te segmenteren. Hoofdstuk 4 behandelt een theoretisch raamwerk om de grijswaardeverdeling in beelden lokaal te beschrijven en laat een aantal mogelijke toepassingen in de beeldverwerking daarvan zien. In hoofdstuk 5 wordt dit raamwerk uitgebouwd tot een textuuranalyse systeem. Een bestaande segmentatiemethode wordt in hoofdstuk 6 met deze textuuraanpak uitgebreid en verbeterd. Ook hoofdstuk 7 draait weer om segmentatie. Hierin wordt een statistische methode beschreven om de ribben in thoraxbeelden te vinden. Tot slot worden in hoofdstuk 8 de verschillende ingrediënten gecombineerd in een systeem dat de longvelden segmenteert en de textuur ervan analyseert. De methode wordt toegepast op beelden van een TB screeningsprogramma en blijkt in staat een groot gedeelte van de abnormaliteiten automatisch te vinden.

Publications

The list contains published as well as submitted articles in international journals and conference proceedings.

- B. van Ginneken, S. Katsuragawa, B.M. ter Haar Romeny, K. Doi, M.A. Viergever, “Local texture analysis in chest radiographs”, Submitted.
- B. van Ginneken, A.F. Frangi, J.J. Staal, B.M. ter Haar Romeny, M.A. Viergever, “Active shape model segmentation with optimal features”, Submitted.
- B. van Ginneken, B.M. ter Haar Romeny, “Statistical local texture analysis applied to computer-aided diagnosis in chest radiography”, in *Statistics of Shapes and Textures — Copenhagen Summer School in Computer Vision*, September 2000.
- B. van Ginneken, B.M. ter Haar Romeny, “Multi-scale texture classification from generalized locally orderless images”, Submitted.
- B. van Ginneken, B.M. ter Haar Romeny, “Automatic delineation of ribs in frontal chest radiographs”, in *Medical Imaging 2000: Image Processing, Proceedings of the SPIE*, vol. 3979, pp. 825–836, 2000.
- B. van Ginneken, B.M. ter Haar Romeny, “Automatic segmentation of lung fields in chest radiographs”, *Medical Physics*, vol. 27, no. 10, pp. 2445–2455, October 2000.
- B. van Ginneken, B.M. ter Haar Romeny, “Applications of locally orderless images”, *Journal of Visual Communication and Image Representation*, vol. 11, no. 2, pp. 196–208, June 2000.
- B. van Ginneken, B.M. ter Haar Romeny, “Automatic segmentation of lung fields in chest radiographs”, in *MICCAI '99 — Medical Image Computing and Computer-Assisted Intervention, Lecture Notes in Computer Science*, no. 1679, pp. 184–191, 1999.
- B. van Ginneken, B.M. ter Haar Romeny, “Applications of locally orderless images”, in *Scale-space theories in computer vision, Lecture Notes in Computer Science*, no. 1682, pp. 10–21, 1999.

- B. van Ginneken, J.J. Koenderink, K.J. Dana, “Texture histograms as a function of irradiation and viewing direction”, *International Journal of Computer Vision*, vol. 31, no. 2/3, pp. 169–184, 1999.
- K.J. Dana, B. van Ginneken, S.K. Nayar, J.J. Koenderink, “Reflectance and texture of real world surfaces”, *ACM Transactions on Graphics*, vol. 18, no. 1, pp. 1–34, 1999.
- B. van Ginneken, M. Stavridi, J.J. Koenderink, “Diffuse and specular reflectance from rough surfaces”, *Applied Optics*, vol. 37, no. 1, pp. 130–139, 1998.
- K.J. Dana, B. van Ginneken, S.K. Nayar, J.J. Koenderink, “Reflectance and texture of real-world surfaces”, in *Proceedings of the CVPR’97*, pp. 151–157, 1997.
- M. Stavridi, B. van Ginneken, J.J. Koenderink, “Surface bidirectional reflection distribution function and the texture of bricks and tiles”, *Applied Optics*, vol. 36, no. 16, pp. 3717– 3725, 1997.
- F. Leymarie, A. de la Fortelle, J.J. Koenderink, A.L.M. Kappers, M. Stavridi, B. van Ginneken, S. Muller, S. Krake, O. Faugeras, L. Robert, C. Gauclin, S. Laveau, C. Zeller, “REALISE: Reconstruction of reality from image sequences”, in *IEEE Proceedings of the ICIP ’96*, vol. III, pp. 651–654, 1996.

Acknowledgements

I had never planned to do a PhD. I found school pretty boring; the university was certainly better, but not every course in the physics curriculum was exciting and the endless series of exams you had to pass were a pain in the ass. But when I could finally do some research myself, I started to enjoy myself more and more. The people I worked with at the department of *Physics of Men* at Utrecht University are largely responsible for this, and I would like to thank them for getting me started in science. So I did a PhD, and it was great. But not just because the subject was challenging and I had almost complete freedom to choose my own directions; also because I worked at the Image Sciences Institute in a very diverse and pleasant environment and got excellent support and input from a great number of people. I'd like to acknowledge those people here.

First of all those who contributed directly to the research described in this thesis. Most important is Bart ter Haar Romeny, my advisor. His positive attitude and enthusiasm are contagious, and I hope we will continue to collaborate in the future. My promotor Max Viergever always gave instant support whenever I walked into his room with questions. With Joes Staal and Alex Frangi I had illuminating discussions about active shape models that certainly left their mark on Chapter 6. Wiro Niessen carefully read the manuscript of Chapter 4. Kunio Doi from the University of Chicago and Shigehiko Katsuragawa of the Nippon Bunri University in Oita, Japan, were my co-workers during my stay in Chicago, when I wrote most of Chapter 8.

This project could never have produced tangible results without the help of radiologists. Several people at UMC Utrecht have been involved. Paul van Waes enthusiastically supported the project from the very beginning and taught Bart and me a lot about chest radiographs, including a one-line definition of an abnormal lung (I'm still working on the implementation). Ad van Gils, Erik Beek, and prof. Puijlaert read a lot of images together with us, thinking out loud. Most of the data I used in this project was obtained from the GGD Midden Brabant in Tilburg. I am extremely grateful to the radiologists Walid Haddad, Joke Annee and Miranda Brouwer who were always interested in my work and enthusiastic about the idea of using computer algorithms to support radiologist in reading screening images. Walid took a lot of time to explain me the varieties of abnormal findings that one encounters on screening images. Only thanks to him I gradually lost that sense of despair that I got whenever I thought about having to write computer programs to detect things that I could hardly see myself. Furthermore, Miranda and Walid drew outlines with a mouse around abnormal areas in over 300 images. In Chicago, Hiroyuki Abe was

“the radiologist in the room next door” (a great luxury to have around). He answered many questions and annotated 100 abnormal chest images for me.

Collecting data is a crucial part of a PhD project in medical image analysis. Ton Froklage of the GGD Midden Brabant processed his complete electronic archive for me. Many of his colleagues helped me with getting the films out from the archive, and back into it. Catelijne Coopmans and Marion van Ginneken also assisted me. Theo van Soest and his colleagues from Radiotherapy at UMC Utrecht helped us with scanning, converting and transferring the images. Marek Brejl, now at Vital Images, and John Haller and Lisann Bolinger from the University of Iowa are gratefully acknowledged for allowing me to use their MR brain data and manual segmentations for Chapter 6.

In the spring of 2000 I visited the Kurt Rossmann Laboratories at the University of Chicago for three months. I am very grateful to Kunio Doi for making this stay possible and arranging many social events. I’d like to thank all members of the Chicago group, especially my room mate Roger Engelmann, for making my visit a great time. During my period in the States, a whole bunch of people made it possible for me to make some cool trips. I thank Hugo de Jong, Steve Pizer, Guido Gerig (visit to University of North Carolina, Chapel Hill), Rik Stokking (Yale University, New Haven), Karel Zuiderveld (Vital Images, Minneapolis and the Mayo Clinic, Rochester), Jimmy Roehrig, Nico Karssemeijer (R2, Los Altos and San Francisco) and everybody I spoke with. Indeed, life of a PhD student isn’t so bad; I also visited conferences and workshops all over the world and spent two weeks with Bart in Rio de Janeiro, one street away from Copacabana Beach. We set up a cooperation with a local TB screening program and were visitors at IMPA, the Instituta Matemática Pura e Aplicada, where Jonas Gomes took great care of us. Hans Haarman, at that time still working for Oldelft, our industrial partner, made the visit to Brazil possible and always showed interest in my project. Later, when Oldelft X-Ray diagnostics became part of Nucletron, his role was taken over by Nol Nuijten.

Writing software was an important part of my work. I’ve learned much from Karel Zuiderveld (and from all the code he left behind when he went to Minneapolis). Theo van Walsum has also been a real guru for me. Without him, it would have taken me weeks more to write what I wrote. In 1999, Theo and I redesigned and rewrote my image processing software environment, imageXplorer, from scratch, a collaboration that I enjoyed very much. imageXplorer is now used throughout our institute. Many other people solved software problems for me and gave me useful pieces of code.

Erik Meijering and Josien Pluim have been my room mates for over three years. They helped me with many things. I enjoyed their company, and I hope they could live with mine, although I know littered our not exactly spacious office with sweaty sport clothes and with food that was often well beyond the state of being edible. Apart from Erik and Josien, there are dozens of people at the Image Sciences Institute and the university who have been a great help and made that I enjoyed my daily work very much. I hope we will continue to have animated discussions, and that we will keep on running, playing soccer, skating and partying together.

Finally, I thank all my friends and family. Especially my parents and, most important of all, Carolin (sorry for the shot noise, but hey, I restored it anyway).



Bram van Ginneken was born in 1970 in Nuenen. At the age of 10, he starred in a school play as Professor Veelweter, a mad scientist who built a machine that produced a three-headed monster named Griezelmans. This role befitted him perfectly: from that moment on he knew what he wanted to be later in life.

After high school at the Augustinianum in Eindhoven, he started to study Physics at the Eindhoven University of Technology, which he interrupted to work as a journalist for some years, and continued at Utrecht University. In 1995 he received a Master of Science degree (cum laude). He stayed as a researcher at the Department of Physics of Man at Utrecht University, until November 1996, when he started his PhD at the Image Sciences Institute, University Medical Center Utrecht. The result of this work is described in this book. He will continue to work at the Image Sciences Institute on computer-aided diagnosis.

Meanwhile, Bram is pursuing a career as a rock-star. In 1986 his first band, The Whistling Haemorrhoids, was founded. Today he is lead singer of Plankton, a noisy combo that has so far released one CD: Pregnant Men Giving Birth to Hell.

<http://www.isi.uu.nl/~bram>

