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](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.
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.