Computer-aided diagnosis in chest radiography: a survey
Ginneken, van, B.; ter Haar Romenij, B.M.; Viergever, M.A.

Published in: IEEE Transactions on Medical Imaging

DOI: 10.1109/42.974918

Published: 01/01/2001

Document Version
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the author's version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Computer-Aided Diagnosis in Chest Radiography: A Survey

Bram van Ginneken*, Bart M. ter Haar Romeny, and Max A. Viergever, Member IEEE

Abstract—The traditional chest radiograph is still ubiquitous in clinical practice, and will likely remain so for quite some time. Yet, its interpretation is notoriously difficult. This explains the continued interest in computer-aided diagnosis for chest radiography. The purpose of this survey is to categorize and briefly review the literature on computer analysis of chest images, which comprises over 150 papers published in the last 30 years. Remaining challenges are indicated and some directions for future research are given.

Index Terms—Chest radiography, computer-aided diagnosis, lung field segmentation, lung nodule detection, rib segmentation.

I. INTRODUCTION

The discovery of X-rays by Wilhelm Conrad Röntgen, in 1895 [1], has revolutionized the field of diagnostic medicine. Today a wide variety of (three-dimensional) imaging techniques is available, and many types of examinations have been, or are about to be replaced by computed tomography (CT) and/or magnetic resonance imaging (MRI) within the next few years. But this is certainly not the case for the chest radiograph. On the contrary, the traditional chest study is still by far the most common type of radiological procedure, making up at least a third of all exams in a typical radiology department [2], [7]. Daffner [2] describes the chest as “the mirror of health and disease.” An enormous amount of information about the condition of the patient can be extracted from a chest film and, therefore, “the ‘routine’ chest radiograph should not be considered quite so routine.”

Interpreting a chest radiograph is extremely challenging. Superimposed anatomical structures make the image complicated. Even experienced radiologists have trouble distinguishing infiltrates from the normal pattern of branching blood vessels in the lung fields, or detecting subtle nodules that indicate lung cancer [4]. When radiologists rate the severity of abnormal findings, large interobserver and even intraobserver differences occur [5], [6]. The clinical importance of chest radiographs, combined with their complicated nature, explains the interest to develop computer algorithms to assist radiologists in reading chest images. The purpose of this survey is to categorize and briefly review the literature on computer analysis of chest radiographs with an emphasis on the techniques that have been employed and on the tasks these techniques are supposed to solve.

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. The first articles about computer analysis of radiographic images appeared in the 1960s [7], [8]. Papers 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 recent review by Duncan and Ayache [9].

These early studies displayed a considerable optimism regarding the capabilities of computers to generate complete diagnoses. They are summarized in one review [10] as attempts to “fully automate the chest exam.” Over the decades this expectation has subsided (which seems to have happened to 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 for decision support rather than trying to make a computer act like a diagnostician [11]; from the abbreviations FACD (fully automatic computer diagnosis), ICD (interactive computer diagnosis), and CAD (computer-aided diagnosis), only the latter is in common use nowadays.

Several related fields are important for CAD in chest radiography, but are outside the scope of this survey because no image processing is involved. Among these are studies on acquisition, about digital chest units versus analog film systems (e.g., [12]–[15]), or on dual energy systems in chest radiography (e.g., [16]–[18]); studies that use psychophysics, e.g., to determine the optimum tube voltage [19] or to aid the detection of abnormalities by measuring visual dwell [20], [21]; attempts to quantify the performance of radiologists, usually through observer studies, when image quality parameters of the chest radiographs are varied (e.g., [22]–[28]); and research on estimating probabilities that patients exhibit a certain disease, given a number of features from clinical information and/or output of computer algorithms. Such studies have continued to appear, starting in the 1960s [29], [30] until recently, e.g., [31]–[33].

Three main areas can be distinguished in the literature on computer analysis of chest radiographs: 1) general processing techniques; 2) algorithms for segmentation of anatomical structures; and 3) analysis aimed at solving a particular task or ap-
plication, usually an attempt to detect a specific kind of abnormality. The following subdivision is adopted and followed in the sequel of this survey.

- 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. [10] reviewed computer analysis of radiographic images in general in 1982, with a strong emphasis on chest radiography. Overviews of research at the Kurt Rossmann Laboratories at the University of Chicago, an active group in computer-aided diagnosis, are given in [34]–[40]. A recent general introduction to computer-aided diagnosis with some examples taken from CAD in chest radiography can be found in [41]. Other informative general overviews can be found in [42] and [43], although the discussion on CAD in chest radiography in both works is strongly focussed on the work done at the University of Chicago. An interesting collection of short papers on CAD can be found in [44]. Reeves and Kostis [45] recently reviewed CAD for lung cancer and although the emphasis is on CT, there is also an overview of methods developed for chest radiographs.

It is worth mentioning here that CAD has been shown to be effective in screening mammography. These successes are certainly helping to make CAD in chest radiography acceptable, notably for the detection of lung nodules. Moreover, the techniques employed in both fields are often similar and may inspire new research. Overviews of CAD in mammography can be found in [42], [46], and [47].

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

Although the skills of experienced radiologists are beyond reach for nonmedical researchers, anyone working on algorithms for the analysis of chest radiographs should obtain some knowledge about the anatomy of the chest, its appearance on projection radiographs and the nature of various abnormal findings. Introductory texts can be found in Squire [48], and in Lange’s standard textbook [49]. A detailed pictorial description of normal and abnormal findings in chest radiographs is given in the chapter on chest imaging in [50]. A comprehensible overview of interstitial disease can be found in [51].

Several tables are included that summarize papers on specific topics. In these tables, studies are listed in chronological order. In most cases information about the evaluation of the presented techniques, such as the size of the test database, are included. We refrained from including any reported figures on classification performance or segmentation accuracy, because these results are usually very dependent on the subtlety of the abnormalities in the test images which makes it impossible to compare the merits of various methods on the basis of such figures alone.

II. GENERAL PROCESSING

A. 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). Such techniques are essential for optimized display of images in soft-copy reading environments that numerous studies have been devoted to this subject. A wide variety of preprocessing procedures for chest radiographs based on local equalization, sharpening, and combinations and modifications of these techniques have been proposed. It is beyond the scope of this survey to give a complete overview. The following works deal with optimal display and image enhancement [52]–[61]. Examples of hardware solutions can be found in [62], [63]. 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.

B. 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 the computer to detect.

A first approach is temporal subtraction, proposed by Kano et al. [64]. An input image is registered with a previous radiograph of the same patient. Elastic matching is employed in which the displacement of small regions of interest (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 stand out as either dark or bright on a gray background. The original technique has been improved and evaluated using subjective ratings by radiologists [65], [66] and the results of the method have been compared with manual registration [67]. An observer study [68] 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. Recently, Zhao et al. [69] proposed a temporal subtraction technique that performs elastic registration on rib border segments detected with an adaptive oriented filter and reported that the contrast around lung nodules substantially increased in the subtracted images. 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. [70], [71] use a scheme similar to Kano et al. [64], based on cross correlation and evaluate the
result with subjective ratings by radiologists. Yoshida [72], [73] minimizes the sum of squared differences, using a mapping that is ensured to be smooth through a regularization term and an optimization in the wavelet domain, and, in [74] segments abnormal areas that “light up” in the subtracted image. The technique is used 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 algorithm to remove ribs by Vogelsang et al. [75], [76] 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.

III. SEGMENTATION

A. 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 [87], the diaphragm [88], or the costophrenic angle (where the diaphragm and the rib cage meet) [89]. A few studies focus on lateral chest radiographs [90], [92].

Table I gives an overview of papers on lung segmentation. The two main approaches are 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 [83], [86]–[92], [98]. Techniques employed are (local) thresholding, region growing, edge detection, ridge detection, morphological 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 [85], [97]). Classifiers are various types of neural networks, or Markov random field modeling, trained with a variety of (local) features including intensity, location, texture measures [84], [85], [93], [94], [97]. Van Ginneken and Ter Haar Romeny [98] combine both approaches in a hybrid scheme.

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 [101]–[105]. In this survey, the term “profiles” refers to an average of consecutive one-dimensional 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. These methods undoubtedly inspired later studies.

Overall, the problem of segmenting lung fields has attracted considerable attention. Several authors report good results that approach the inter-observer variability ([98] lists a comparison of quantitative results of several studies). The task of segmenting lung fields might, therefore, 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% 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 a worthwhile endeavor.

### Table I

An overview of literature 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.

<table>
<thead>
<tr>
<th>Method</th>
<th>RB</th>
<th>PC</th>
<th>PA/lateral</th>
<th>evaluation</th>
<th>remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toriwaki [77], [78]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>describes a complete analysis system</td>
</tr>
<tr>
<td>Harlow [79]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chien [80], [81]</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>only right lung, used to detect abnormalities</td>
</tr>
<tr>
<td>Hasegawa [82]</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pietka [83]</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McNitt-Gray [84], [85]</td>
<td>X</td>
<td></td>
<td></td>
<td>16 Q</td>
<td>uses 5 anatomical classes</td>
</tr>
<tr>
<td>Duryea [86]</td>
<td>X</td>
<td></td>
<td></td>
<td>802 Q</td>
<td></td>
</tr>
<tr>
<td>Xu [87]</td>
<td>X</td>
<td></td>
<td></td>
<td>1000 R</td>
<td>outer rib cage only</td>
</tr>
<tr>
<td>Xu [88]</td>
<td>X</td>
<td></td>
<td></td>
<td>300 R</td>
<td>diaphragm edges only</td>
</tr>
<tr>
<td>Armato [89]</td>
<td>X</td>
<td></td>
<td></td>
<td>600 R</td>
<td>costophrenic angles only</td>
</tr>
<tr>
<td>Armato [90]</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Armato [91]</td>
<td>X</td>
<td></td>
<td></td>
<td>600 R</td>
<td></td>
</tr>
<tr>
<td>Carrascal [92]</td>
<td>X</td>
<td></td>
<td></td>
<td>65 RQ</td>
<td></td>
</tr>
<tr>
<td>Vittitoe [93]</td>
<td>X</td>
<td></td>
<td></td>
<td>99 Q</td>
<td></td>
</tr>
<tr>
<td>Tsuji [94]</td>
<td>X</td>
<td></td>
<td></td>
<td>71 Q</td>
<td></td>
</tr>
<tr>
<td>Wilson, Brown [95], [96]</td>
<td>X</td>
<td></td>
<td></td>
<td>115 Q</td>
<td>uses 6 anatomical classes</td>
</tr>
<tr>
<td>Vittitoe [97]</td>
<td>X</td>
<td></td>
<td></td>
<td>115 Q</td>
<td></td>
</tr>
<tr>
<td>Van Ginneken [98]</td>
<td>X</td>
<td></td>
<td></td>
<td>115 Q</td>
<td>general knowledge-based method</td>
</tr>
<tr>
<td>Van Ginneken [99]</td>
<td>X</td>
<td></td>
<td></td>
<td>115 Q</td>
<td></td>
</tr>
</tbody>
</table>
Finally, we note that it is often hard to establish ground truth data for evaluation of segmentation techniques. This problem applies to all segmentation tasks discussed here. Preferably, the objects should be delineated by several experienced observers. In the case of lung fields, the definition of a lung field is already unclear, because large parts of the lungs are obscured in PA chest radiographs [106]. In rib segmentation, the overlapping posterior and anterior rib parts make a definition of “rib borders” difficult.

### B. 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, ROIs between (the intercostal space) or on (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 II. Several methods also detect (parts of) the contours of the lung fields [111], [112], [114], but an evaluation of their accuracy is not included. The classical approach to rib segmentation, used in most studies, is as follows. A geometrical model of rib borders is chosen, e.g., parabolas or ellipses or a combination. Edge detection extracts segments that may be parts of the rib borders. These candidate rib border parts are analyzed and rejected, or grouped together into complete ribs. In other studies, each candidate votes for combinations of model parameters (a modified Hough transform). Postprocessing steps may remove some borders, or infer new rib borders that had not been detected. A fine-tuning stage may be added in which the strict geometric model is abandoned, such as the snakes fitted to rib borders in [114].

A difficulty in the classical approach is that the relations between ribs are not 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 [110] do not appear very powerful or use simple ad hoc rules [114]. This was already recognized by Wechsler [109] 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 [115] is an attempt to model the shape of the complete rib cage and fit it directly to the image.

The number of images used in evaluation is in general considerably smaller than in the case of lung field segmentation 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 II). Since the full rib cage has a much more complex shape than the outline of the lung fields, partial failures occur frequently. 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.

C. Other Structures

Toriwaki et al. describe a complete system for the analysis of chest radiographs [77], [78] that includes segmentation of lung fields, rib cage, heart, clavicles, and blood vessels followed by automatic detection of abnormalities. Only a rough description of results on 15 [77] and 40 images [78] taken from a mass screening for lung tuberculosis in Japan is given without mention of the nature and subtlety of the abnormalities in this database.

A method to select small ROIs within the lung fields that do not contain rib borders is presented by Chen et al. [116]. The lateral lung fields are subdivided into regions which are partly eliminated if they contain edges of certain strength and orientation.

Several methods for segmentation of (parts of) the heart boundary have been proposed [8], [76], [101], [117]–[120], usually with the aim of proposed [8], [76], [101], [117]–[120], 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 (CTR) indicative of cardiomegaly, since parts of the boundaries of the lung fields coincide with the heart contour. In Section IV-A, these methods will be discussed in some more detail.

Armato et al. [121] 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.

IV. ANALYSIS

What kind of abnormalities does the radiologist who reads chest radiographs encounter in his/her day-to-day practice? The answer to this question should set the agenda for computer scientists who want to—as it was put in Section I—“fully automate the chest exam.”. MacMahon [122] 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 of this study are summarized in Table III. The statistics are surprising when compared with abnormalities commonly investigated in the literature on computer analysis of chest images. Lung nodules are a relatively rare abnormality but have received much attention in the 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 analyses and it is important to realize how common such findings are. The same is true for clothing artifacts, that are present in many chest radiographs.

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

- Lung nodule detection.

- Detection of cardiomegaly or estimation of the CTR.
- 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.”

A. Size Measurements

In a few applications, diagnostic information can be extracted directly from segmentations. An example is the estimation of enlarged heart size (cardiomegaly) usually performed by measuring the CTR, the maximal horizontal diameter of the heart divided by the maximal horizontal diameter of the thorax. This is the subject of probably the earliest computer analysis studies of chest radiographs by Becker, Meyers, and co-workers [7], [8]. Points on the rib-cage boundary and heart boundary are extracted from horizontal profiles and the CTR is estimated for 37 radiographs. In work by Hall [117] and Kruger [101], and their co-workers, several parts of the lung and heart boundaries were segmented. Images were classified in one of three categories (normal, rheumatic heart disease, or otherwise abnormal) using various ratios of distances derived from the segmentation as

<table>
<thead>
<tr>
<th>Finding</th>
<th>% of all abnormals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pulmonary infiltrates</td>
<td>55%</td>
</tr>
<tr>
<td>Intravenous catheters</td>
<td>33%</td>
</tr>
<tr>
<td>Heart size/contour</td>
<td>27%</td>
</tr>
<tr>
<td>Endotracheal/tracheostomy tubes</td>
<td>22%</td>
</tr>
<tr>
<td>Pleural effusions</td>
<td>12%</td>
</tr>
<tr>
<td>Linear atelectasis/scar</td>
<td>10%</td>
</tr>
<tr>
<td>Drainage catheters and tubes</td>
<td>9%</td>
</tr>
<tr>
<td>Pulmonary vascularity</td>
<td>9%</td>
</tr>
<tr>
<td>Pleural scarring</td>
<td>8%</td>
</tr>
<tr>
<td>Rib lesions</td>
<td>7%</td>
</tr>
<tr>
<td>Mediastinal masses</td>
<td>6%</td>
</tr>
<tr>
<td>Diaphragm</td>
<td>5%</td>
</tr>
<tr>
<td>Calcified granulomas</td>
<td>5%</td>
</tr>
<tr>
<td>Pneumothorax</td>
<td>5%</td>
</tr>
<tr>
<td>Lung nodules</td>
<td>5%</td>
</tr>
<tr>
<td>Extrathoracic abnormalities</td>
<td>4%</td>
</tr>
<tr>
<td>Lung masses</td>
<td>2%</td>
</tr>
<tr>
<td>Calcified nodules</td>
<td>1%</td>
</tr>
<tr>
<td>Mediastinal shift/contour</td>
<td>1%</td>
</tr>
<tr>
<td>Cardiac pacemakers</td>
<td>1%</td>
</tr>
</tbody>
</table>
features in linear and quadratic discriminant functions. The method was evaluated on 320 films, a surprisingly large number for studies from the 1970s. Using a similar approach, Sezaki and Ukena [118] computed the CTR by 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. [123] used the method from [101] to detect points on the heart boundary to which a Fourier shape was fitted. The coefficients of the Fourier shape were used to classify the heart shape as normal or abnormal, giving reasonable results on a database of 481 images with 209 normal cases. A comparable method was proposed by Nakamori et al. [119]. A Fourier shape was fitted to points on the heart boundary found by edge detection and the results were used to detect cardiomegaly in [120] and evaluated on a database of 400 radiographs with 91 abnormalities. In conclusion, computer estimation of the CTR has received considerable attention and promising results have been obtained. 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 limited number of 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 decade.

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 an empiric formula that assumes simple shapes for the lungs and that has been shown to correlate well with the true TLC. Paul et al. [103] described a method to automatically determine the TLC using profile analysis similar to Kruger et al. [101]. It was tested on 15 radiographs. Carrascal et al. [92] used a rule-based segmentation of PA and lateral lung fields to estimate the TLC for 65 radiographs.

B. Nodule Detection

Automatic detection of lung nodules is the most studied problem in computer analysis of chest radiographs. One in every 18 woman and every 12 men develop lung cancer, making it the leading cause of cancer deaths. Early detection of lung tumors (visible on the chest film as nodules) may increase the patient’s chance of survival. But detecting nodules is a complicated task; see, e.g., [4]. In a lung cancer screening program for heavy smokers in which chest radiographs were taken every four months, it was shown that for 90% of peripheral lung cancers that were detected, nodules were visible on earlier images, when these older images were checked retrospectively [159].

Nodules show up as relatively low-contrast white circular objects within the lung fields. The difficulty for CAD schemes is to distinguish true nodules from (overlapping) shadows from vessels and ribs.

Almost all methods rely on a two-step approach for nodule detection. In the first stage initial candidate nodules are detected. The second stage consists of eliminating as many false positive candidates as possible, without sacrificing too many true positives. Table IV lists methods for the detection of candidate nod-
TABLE V
AN OVERVIEW OF METHODS FOR CLASSIFYING 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 FOR EVALUATION (THIS IS NOT NECESSARILY THE NUMBER OF IMAGES AND NODULES USED FOR VALIDATION ONLY)

<table>
<thead>
<tr>
<th>Study</th>
<th>Features</th>
<th>Classifier</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ballard [125], [126]</td>
<td>Radius, contrast, distance from hilum, shape features.</td>
<td>nearest neighbor thresholds</td>
<td>6 images, 8 images, 30 nodules</td>
</tr>
<tr>
<td>Sankar [137]</td>
<td>Shape and contrast features from boundary obtained with dynamic programming.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sklansky [127]</td>
<td>13 features based on histograms and output of filters applied to ROIs.</td>
<td>thresholds</td>
<td>19 images, 16 abnormals</td>
</tr>
<tr>
<td>Lamper [138], [128]</td>
<td>Rib test (presence of parallel edges), spatial location measures, Hough transform response, radius of circle, gray level features.</td>
<td>linear discriminant thresholds</td>
<td>37 abnormals</td>
</tr>
<tr>
<td>Giger [129]</td>
<td>Circularity, size and growth rate of region grown area for various thresholds.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Giger [189]</td>
<td>Circularity of region grown area for various thresholds after morphological opening.</td>
<td>threshold</td>
<td>60 images, 33 abnormals</td>
</tr>
<tr>
<td>Cox [140]</td>
<td>Horalick's [141] and Laws' [142] texture features obtained from ROIs with and without nodules.</td>
<td>ANN, ROIs from 5 images</td>
<td></td>
</tr>
<tr>
<td>Matsumoto [143]</td>
<td>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.</td>
<td>thresholds</td>
<td>60 images, 30 abnormals, 32 nodules</td>
</tr>
<tr>
<td>Suzuki [131], [132]</td>
<td>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.</td>
<td>thresholds, linear discriminate, feature selection</td>
<td>192 nodules</td>
</tr>
<tr>
<td>Peter Chiou [144]</td>
<td>Single scale edge strength and orientation.</td>
<td>Kohonen map</td>
<td>31 images, all abnormal, 87 nodules</td>
</tr>
<tr>
<td>Wu [145]</td>
<td>Similar to [145].</td>
<td>ANN, discriminant analysis, thresholds, convolution neural net</td>
<td>60 images, 30 abnormal, 32 nodules</td>
</tr>
<tr>
<td>Lo [146]</td>
<td>$32 \times 32$ raw image data from background corrected candidate ROIs.</td>
<td></td>
<td>55 images, 25 abnormal, 52 nodules</td>
</tr>
<tr>
<td>Lin [147], [148]</td>
<td>$32 \times 32$ raw image data from background corrected candidate ROIs.</td>
<td>two convolution neural nets, threshold</td>
<td>In [148]: 54 images, 23 abnormals</td>
</tr>
<tr>
<td>Floyd [149]</td>
<td>Fractal dimension estimated from power spectrum.</td>
<td></td>
<td>25 images, 10 abnormal</td>
</tr>
<tr>
<td>Vittitoe [150]</td>
<td>Fractal dimension estimated from power spectrum.</td>
<td>threshold</td>
<td>30 images, real and simulated nodules</td>
</tr>
<tr>
<td>Xu [151]</td>
<td>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.</td>
<td>threshold, followed by ANN</td>
<td>200 images, 100 abnormal, 122 nodules</td>
</tr>
<tr>
<td>Carreira [152]</td>
<td>Size as a function of thresholds, various shape features from the intensity landscape.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Penedo [153]</td>
<td>Curvature of intensity landscape.</td>
<td>ANN</td>
<td>35 abnormal images, 70 nodules</td>
</tr>
<tr>
<td>Xu [154]</td>
<td>Mean, minimum, maximum, width, standard deviation of histograms of gradients perpendicular to candidate boundaries for various segments of the boundary.</td>
<td>thresholds</td>
<td>60 images, 90 nodules, 288 simulated nodules</td>
</tr>
<tr>
<td>Casalidi [155]</td>
<td>Morphological operations are used to detect likely location of nodules.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nakamura [156]</td>
<td>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.</td>
<td>ANN classifies nodules as benign / malignant</td>
<td>56 cases, 34 benign, 22 malignanat</td>
</tr>
<tr>
<td>Yoshida [73]</td>
<td>After local contralateral subtraction to remove rib structures, signal to noise ratio in candidate ROIs.</td>
<td>threshold</td>
<td>550 candidates, 51 nodules</td>
</tr>
<tr>
<td>Catarious [157]</td>
<td>Channelled Hotelling statistics, fractal dimension, geometric features Template matching with nodule and non-nodule candidates.</td>
<td>ANN</td>
<td>237 candidates, 228 images, 250 nodules</td>
</tr>
</tbody>
</table>

Table V lists schemes for eliminating false positive candidates. Several methods include both steps but focus on one of them in particular. These methods are included in the most appropriate table.

Several schemes start with producing an image in which nodules are enhanced. This is done by filtering with a nodule-like filter and/or suppressing background structures by some sort of blurring [129], [133], or by applying preprocessing techniques such as unsharp masking or similar operations [138]. Nodule candidates are detected using template matching or a modified Hough transform in which edge pixels vote for circles that could cause these edges [125], [126], [128], [135], [138]. In other cases the nodule enhanced image is simply thresholded. Some studies apply a background trend correction on ROIs with nodules candidates [143], [154], which is similar to high-pass filtering of the input image as a preprocessing step.

The next step consists of eliminating false positive responses. Usually specific features are detected for each nodule candi-
date. Table V gives a brief description of features used in each method. Almost all features are (variations on) “classical features” in the sense that they can be found in standard image processing textbooks, see for instance [160, Chapter 9]. 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 [146], [148].

It is clear from Table IV 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 on this problem is limited). Judging the relative merits of methods on the basis of these papers 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. An exception, and probably the first attempt to set up a common database for lung nodules is the work by Shiraishi and co-workers [161], who collected a database of 247 chest radiographs with 154 nodules that is available on CD-ROM. 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 crucial to progress in this field. Unless these databases are publicly available, it will remain difficult to compare performance of different methods (see also [162]). The evaluation is an important part of studies on computerized nodule detection. It is encouraging that receiver operating characteristic methodology [163], [164] has become the standard tool for evaluation. In general, nodule detection remains an open and difficult task in computer-aided diagnosis that is still far from solved.

A related problem is the classification of nodules as benign or malignant. Nakamura et al. [156] describe a method in which a feed-forward neural network outperformed radiologists in classifying nodules as benign/malignant based on features computed from manually drawn outlines of the nodules. Gurney and Swensen [165] 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). A preliminary conclusion, based on these two studies, could be that even if the performance of a computer in detecting nodules still lags that of radiologists, computer algorithms may be useful to enlarge existing nodule databases. An example is the work of Sherrier et al. [171], who construct simulated nodules images by identifying isolated nodules, without any rib structures nearby, and adding rotated and scaled versions of these nodule templates to new radiographs, at various locations. Physically there is not much wrong with this procedure since radiographs are projection images, although it could be doubted whether the distribution of size and structure of nodules does not depend on the location in the lung field, and the scatter contributions of nodules of different size and orientation would be different. Observer studies in [171] showed that radiologists could not discern between simulated and true nodule cases.

C. Texture Analysis

Many diseases of the lungs are characterized by diffuse patterns in chest radiographs. These diseases are often referred to as diffuse lung disease or interstitial disease (although strictly speaking the latter is a smaller category). The interstitium of the lung is the connective tissue between the blood vessels and the alveoli, the tiny air sacs. Sorting out interstitial disease is one of the most difficult tasks for a chest radiologist [51]. Clues are the type of patterns (linear, reticular, nodular, honeycomb), the location, shape, and symmetry of the affected areas, the borders of these areas (well- or ill-defined) and their changes over time. From an image processing point of view, texture analysis is the proper way to analyze these kinds of abnormalities, hence, the title of this section.

The difference between computer detection of nodules and that of interstitial abnormalities is that interstitial abnormalities have a more diffuse character and, therefore, the two-stage approach for lung nodule detection (finding candidates and eliminating false positives) is not easy to apply. Instead, all areas within the lung fields should be checked for the presence of patterns of interstitial disease.

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 [172]. Revesz and Kundel investigated the feasibility of classification based on features computed from the (optically determined!) Fourier spectrum [173]. A similar device was used later by Stark and Lee [174] who compared the performance of several classifiers with a database of 64 chest films. Kruger, Turner, and Thompson [175], [176] computed features derived from co-occurrence matrices (cf. [141]) 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 [102]. Later work by Hall [104], [177], used a coarse method for automatic segmentation of the lung fields and presented the design for a complete system. Ledley et al. [178] 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 [105], [179], [180]. Li et al. [181] treated the problem as one of detecting small rounded opacities and employed a strategy comparable with the standard nodule detection strategy. Classification is based on contrast features computed from region grown areas. Recently, Soliz et al. [182] presented a system for the detection of pneumoconiosis using a specific type of neural network and features derived from co-occurrence matrices from manually selected ROIs. 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 [64]: 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. [183], [184] 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 [116], and computation of the standard deviation and the first moment of the Fourier spectrum [185], [186]. These two features are used to distinguish three classes of interstitial disease patterns (nodular, reticular, and honeycomb) from normal tissue. The same method has been applied to the standard radiographs from the International Labour Office for detection of pneumoconiosis in [187]. In other studies, geometric features from edges and ridges (line filters) and blobs (found by thresholding) were investigated [188], [189] and improved [190], and horizontal profiles of ROIs were used as input to a feed-forward neural network [191]. 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 patterns [191], [192]. The use of the scheme was also tested in an observer study [193].

A few studies on the detection of interstitial disease have been done by other groups. Kido et al. [194] computed geometric features, similar to the method later described in [189]. In another work, Kido et al. [195] 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 [196], [197] and try to correct for these effects [198].

In [99], 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 with the results reported in [191]. The performance on the tuberculosis database was substantially lower. From the difference in performance of the same scheme on two databases we conclude once again that it is hard to compare results from different studies if different image sets are used. There are no established common databases to test the performance of computer algorithms for detecting interstitial disease. The results published so far seem to indicate that the detection of subtle interstitial abnormalities is still an open problem.

D. 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. [199] 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 [122]. Vogelsang et al. [76] present a method to detect catheters but do not present an evaluation.

V. DISCUSSION AND RECOMMENDATIONS FOR FUTURE RESEARCH

This survey gave an overview of three decades of research on computer analysis of chest radiographs in which over 150 papers on the subject have appeared. In this final section, we summarize the results by listing a number of solved and unsolved problems and possibilities for clinical applications by identifying some trends in the field, and with a list of recommendations for future research.

A. Solved and Unsolved Problems and Possibilities for Clinical Applications

• Several methods for segmentation of lung fields have obtained good results on test databases of reasonable size.
• Segmentation the heart and other large structures, such as the clavicles, has not received so much attention, but seems feasible as well.
• Automatic delineation of the posterior and anterior ribs is a harder problem for which no thoroughly evaluated methods have been proposed yet.
• Subtraction techniques (filtering out normal structures) have shown impressive results in distinct cases but are not yet applied in clinical practice.
• Most research has been focused at the detecting of lung nodules, and this has resulted in a number of schemes that have been evaluated on reasonably sized databases. For better performance in subtle cases, progress is needed.
• The results of texture analysis in chest radiograph are encouraging, but progress is needed to detect more subtle cases.
• Segmentation of branching blood vessels, and detection of objects such as clothing and catheters has not received much attention, although the results of such analyses could be used to eliminate false positives, to choose ROIs for texture analysis and to subtract normal structures. This remains an open problem.

The automatic detection of enlarged heart size and the total lung volume seem likely candidates for clinical application. Subtraction techniques seem also clinically applicable, especially for the detection of interval change.

However, lung nodule detection is the first area in which industry has developed computer-aided diagnosis products. This problem is difficult and far from solved, but the consequences of missed lung nodules spur several industries to develop products to assist radiologist in this task. Missed lung cancer is the second most common cause of malpractice suits among radiologist in the United States. In 90% of these cases the alleged error occurred on chest radiographs. About half of these cases resulted in indemnity payments averaging $150 000 per case [200]. This year the first product for lung nodule detection, RapidScreen by Deus Technologies, Rockville, MD, has received U.S. Food and Drug Administration marketing approval.

B. Trends

• 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.

• Radiology departments are starting to use central systems to store digital chest radiographs. This provides an opportunity to use “chest workstations” with intelligent algorithms that run in the background and may alert radiologists to possible abnormalities.

• 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. Computer algorithms may become an important aid to the radiologist in differential diagnosis.

• We expect to see more applications of statistical knowledge-based techniques. Such methods are gaining popularity in many areas of medical image processing and the analysis of chest radiographs, where large numbers of images are available, seems ideally suited for their application.

C. Recommendations

• 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 (see, [122] and Table III). The clinical experience with chest workstations should set the agenda for further research. CAD schemes should be developed and tested with radiographs as they are encountered in clinical practice. This is in contrast with previous research, in which the focus always was on a single aspect or abnormality to be detected, and where algorithms were tested on a selected database with only those abnormalities and normal cases.

• Segmentation methods should preferably include quantitative results, offset against intraobserver and interobserver variability. The value of schemes that detect abnormalities can only be assessed when they are compared in observer studies with the performance of radiologists, or by comparing the performance of radiologists with and without the aid of a computer. In the latter case, the task for the radiologist should be similar to clinical practice.

• The use of dual energy subtraction images in computer analysis, that can be produced virtually for free by several commercially available chest units, could improve the sensitivity of CAD systems.

• Larger, publicly available databases should be collected and used for better validation. The work of Ho and Kruger [161] serves as an example. Research programs and funding agencies should become aware of this issue. These databases will reduce the burden of individual research groups to collect data and allows fair comparisons of different techniques. So far, many papers report good or excellent results, but this can be deceiving because abnormalities in chest radiographs occurs over such a wide range of subtleties. Increasing the size of databases is not only important for validation, but also for the development of better CAD schemes, that use data for training. Often the number of actually positive cases is limited compared with the number of features used for classification or feature selection or the number of parameters used in rule-based schemes.

• To increase the size of databases, it can be useful to explore the use of methods that use existing data to simulate new cases.

• Comparative studies are needed to assess the value of different algorithms. Researchers should be urged to make code implementations publicly available, to facilitate such studies.

ACKNOWLEDGMENT

This paper is a result of a four-year project on computer-aided diagnosis in chest radiography performed at the Image Sciences Institute, University Medical Center Utrecht, The Netherlands, which was supported by the Dutch Ministry of Economic Affairs through the IOP Image Processing program.

REFERENCES


