Automatic Detection of Abnormalities in Chest Radiographs Using Local Texture Analysis

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

***Abstract*—A fully automatic method is presented to detect abnormalities in frontal chest radiographs which are aggregated into an overall abnormality score. The method is aimed at finding abnormal signs of a diffuse textural nature, such as they are encountered in mass chest screening against tuberculosis (TB). 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 multiscale 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 receiver operating characteristic (ROC) curve of 0.820. The second database consist 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.**

***Index Terms*—Chest radiographs, computer-aided diagnosis, texture analysis.**

I. Introduction

DESPITE the existence of a cheap cure, tuberculosis (TB) is a leading killer of adults in the world [1]. 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 (CAD) 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 paper, we propose a scheme to detect such textural abnormalities.

The important role of texture analysis in CAD is increasingly recognized [2]. 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 been studied extensively. Early work dates from the 1970s and focuses on the detection of pneumoconiosis, using features extracted from pixel profiles, Fourier spectra and cooccurrence matrices [3]–[10]. More recent methods are applied to interstitial disease in general, and essentially use a similar approach: regions of interest (ROIs) are (automatically [11]) selected and texture features are computed, based on the Fourier spectrum [12]–[15], geometric features [16], responses to filters [17], pixel profiles [18], fractal dimension [19], or combinations of feature sets [20]. 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 [21]–[23]. Filter bank methods, in which statistics from filtered images are used as features, have been shown to be very effective, if used within a multiresolution framework [24]–[28]. This motivates our choice to use moments of histograms extracted from regions in the image after filtering with a multiscale filter bank consisting of the Gaussian 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 ROIs that do not contain normal structures such as (crossing) rib borders and (large) vessel projections [11], [13]. An alternative could be to preprocess the images so as to remove normal background structures [29], [30], 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 [12]. 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

Manuscript received January 17, 2001; revised December 14, 2001. The work of B. van Ginneken was supported by the Dutch Ministry of Economic Affairs through the IOP Image Processing program, which made it possible to visit the Kurt Rossmann Laboratories at the University of Chicago, during which period most of this study was carried out. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was M. Sonka. *Asterisk indicates corresponding author.*

*B. van Ginneken is with the Image Sciences Institute, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands (e-mail: bram@isi.uu.nl).

S. Katsuragawa is with NBU General Research Center, Nippon Bunri University, 870-0397-Oita, Japan.

B. M. ter Haar Romeny is with the Department of Medical and Biomedical Imaging, Faculty of Biomedical Engineering, Technical University Eindhoven, 5600 MB Eindhoven, The Netherlands.

K. Doi is with the Kurt Rossmann Laboratories for Radiological Image Research, Department of Radiology, University of Chicago, Chicago, IL 60637 USA.

M. A. Viergever are with the Image Sciences Institute, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands

Publisher Item Identifier S 0278-0062(02)02942-7.

at multiple scales, but also from multiple "apertures." Thus, we perform *local* texture analysis with multiple *inner* scales and *outer* scales.

II. Materials

A. 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 abnormals—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 Electrodelca units (Oldelft BV, Delft, 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 areas 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.

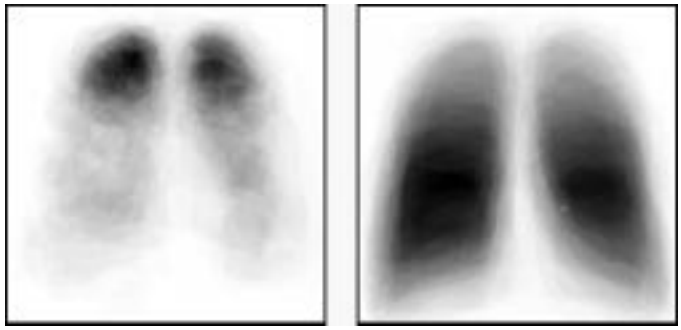
The ground truth that we are trying to reproduce by computer for this database is the judgment of the radiologists who read the films. In a large number of cases (106 out of the 279 abnormal cases), one of the radiologists judged the image to be normal, while the other favored to contact the patient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

TB is known to often affect the upper lung fields. Fig. 1(a) gives an idea of the spatial distribution of the abnormal areas within the chest radiographs in the TB database.

For this study, all normal cases and only those abnormal cases with textural abnormalities were used. Furthermore, all images with large clothing artifacts (people are photographed fully dressed in this screening) 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.

B. 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 from daily clinical practice. 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 $\times$2000 pixels with 0.175—mm pixel size and 10-bits intensity.



(a) (b)

Fig. 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.

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. Fig. 1(b) gives an idea of the spatial distribution of the abnormal areas within the chest radiographs.

III. Methods

A. Segmentation

1) Active Shape Model (ASM) 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. Rule-based schemes have been proposed by Armato *et al.* [31], Xu *et al.* [32], [33], Duryea and Boone [34], Pietka [35], and Brown *et al.* [36]. Lung segmentation by pixel classification using neural networks has been investigated by McNitt-Gray *et al.* [37] and Tsujii *et al.* [38]. Vittitoe *et al.* [39], [40] 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 [41].

In this paper, the lung fields are delineated with ASMs, a general technique for image segmentation. ASMs have been developed by Cootes and Taylor [42] and have been applied to various segmentation tasks in medical imaging [43], [44]. Below the ASMs scheme is described briefly. For details about the method, see [42] or [45].

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 [46] as follows.

The landmark points $(x_1, y_1), \ldots (x_n, y_n)$ are stacked in shape vectors

$$\mathbf{x} = (x_1, y_1, \ldots, x_n, y_n)^T. \tag{1}$$

Given a set of s shapes, Principal Component Analysis is applied to these vectors (without performing any kind of alignment), by computing the mean shape and the covariance

$$\overline{\mathbf{x}} = \frac{1}{s}\sum_{i=1}^{s}\mathbf{x}_i \tag{2}$$

$$\mathbf{S} = \frac{1}{s-1}\sum_{i=1}^{s}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T \tag{3}$$

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

$$\mathbf{x} \approx \overline{\mathbf{x}} + \mathbf{\Phi}\mathbf{b} \tag{4}$$

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

$$\mathbf{b} = \mathbf{\Phi}^T(\mathbf{x} - \overline{\mathbf{x}}). \tag{5}$$

When fitting the model to a set of points, the values of $\mathbf{b}$ are constrained to lie in the range of several 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

$$\sum_{i=1}^{t}\lambda_i \geq f_v\sum_{i=1}^{2n}\lambda_i. \tag{6}$$

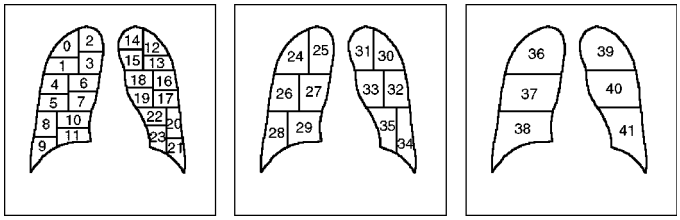
To create models of the image profiles around each landmark, profiles $\mathbf{g}_1 \cdots \mathbf{g}_s$ are sampled (using linear interpolation) around each landmark, perpendicular to the contour. Sampling k pixels on either side of the profiles gives profiles of length $2k+1$. The first derivatives 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 $\overline{\mathbf{g}}$ and the covariance matrix $\mathbf{S}_g$ are computed. To fit the model, the Mahalanobis distance between a new profile $\mathbf{g}_i$ and the profile model can be computed

$$f(\mathbf{g}_i) = (\mathbf{g}_i - \overline{\mathbf{g}})\mathbf{S}_g^{-1}(\mathbf{g}_i - \overline{\mathbf{g}}). \tag{7}$$

Minimizing this Mahalanobis distance $f(\mathbf{g}_i)$ is equivalent to maximizing the probability that $\mathbf{g_i}$ originates from the distribution $\{\mathbf{g}_1 \cdots \mathbf{g}_s\}$. This minimization is used to find new locations for the landmarks during fitting.

These profile models, given by $\overline{\mathbf{g}}$ 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. Note that we do not, as proposed by Cootes, subsample the images.

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.



(a) (b) (c)

Fig. 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 in subsequent tables. The shape and location of the lung fields shown here are obtained from the mean lung fields of the manually segmented training images.

2) 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 [Fig. 2(c)]. These parts are subdivided into four areas of equal area, by first subdividing them vertically [Fig. 2(b)] and then horizontally [Fig. 2(a)]. In this way, a total of 42 regions are defined.

After segmentation, the region maps as shown in Fig. 2 are warped to the segmentation result, using interpolation with radial basis functions [47]. After warping, the borders between the regions are no longer necessarily horizontal or vertical. 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.

B. Texture Analysis

We use a multiscale 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 rationale behind the choice for this set of filters is that they describe local image structure; they are the coefficients of a Taylor expansion of the image [48]. The choice of a specific filter bank is important, but not crucial. For instance, Gaussian derivatives are very similar to Gabor functions, and similar results can be obtained with an appropriate set of Gabor filters. The exact choice of orientation and scale of filters was determined empirically.

The images in the TB database were subsampled to a working resolution of 700 × 700 pixels and six 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 four scales $\sigma = 1, 2, 4, 8$ pixels. The images in the ID database were subsampled to 1400 × 1400 pixels 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. Note, however, that there are structural differences between the two sides (the right lung has three lobes, and the left lung two, and the heart is on the left side). The difference features may be capturing the structural and not the textural differences.

C. 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 (NNs) 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 NNs developed by Arya and Mount [49], 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 NNs. If there are k neighbors at distances $d_1, \ldots d_k$ from the feature vector to be classified, and $c_1, \ldots, c_k$ 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

$$c = \frac{\sum_{i=1}^{k} \exp\left(\frac{-d_i^2}{\sigma^2}\right) c_i}{\sum_{i=1}^{k} \exp\left(\frac{-d_i^2}{\sigma^2}\right)} \tag{8}$$

where σ is the mean of $d_1, \ldots, d_k$.

The classification is a number in the range from zero (normal) to one (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 [50], [51]. 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. This is done by fitting a model to the data that assumes that the ROC curve has the same functional form as one produced by a pair of two normal underlying distributions. This bi-normal model is fitted using maximum-likelihood estimation. The software we used to fit the ROC curves is freely available at http://www-radiology.uchicago.edu/sections/roc.

D. 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 paper, 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 region 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

$$c = 1 - \prod_{i=1}^{N} (1 - f_i c_i) \tag{9}$$

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

$$f_i = \mathrm{Max}\left(\frac{A_{zi} - T_{A_z}}{1 - T_{A_z}}, 0\right). \tag{10}$$

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

IV. Results

A. Segmentation

To train the ASM, 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 × 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 six 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 five 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 Figs. 4–6, the warped regions after segmentation for several normal and abnormal images from the TB database are shown. Table I 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 images 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 II. A failure is defined as a segmentation in which at least one of the 24 smallest regions into which the lung fields are divided (Fig. 2) is 50% incorrect.

B. Texture Analysis and Classification

For the classification of texture vectors computed from regions, a value of k NNs 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 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 III.

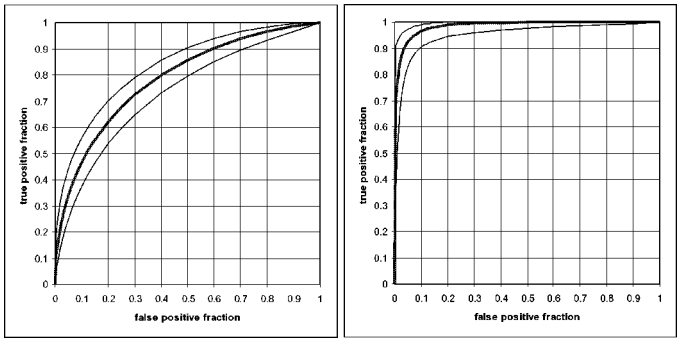


Fig. 3. ROC curves for the best performing feature sets. The thin lines below and above the curve denote the asymmetric 95% confidence intervals. (a) TB database, features set $\sigma = 2$; only second moment and only difference features. The area under the curve is 0.820. (b) ID database, feature set σ = 1, 2, 4; second, third, and fourth moments. The area under the curve is 0.986.

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 $\pm$ 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 one, the perfect result. Because the resolution of the ID images is higher, we used one more scale in the multiresolution 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. We do not have an explanation for these differences. The best result, $A_z = 0.986 \pm 0.0062$, is obtained using the original features for three scales σ =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 $A_z = 0.98$.

Fig. 3 plots the ROC curves for the best performing features sets for the TB database and the ID database. Table IV lists the A_z values for each region for the TB database and the ID database, using the best performing feature sets.

V. Discussion

It is interesting to note that the overall performance of $A_z =$ 0.820 on the TB database is much higher than the A_z values of

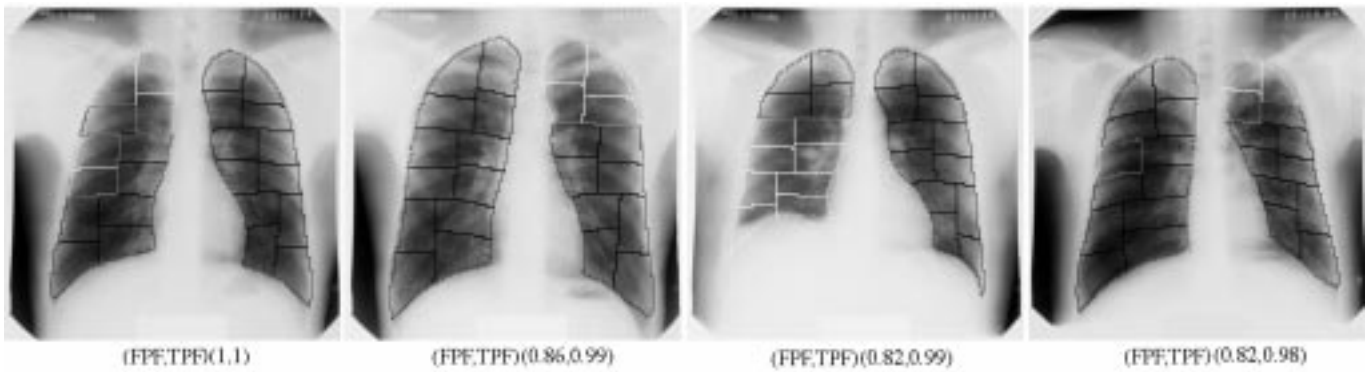


Fig. 4. Of all abnormal cases in the TB database, the four images shown here were judged most normal according to the computerized scheme. These are the most difficult positive cases for the CAD scheme. Below each image its position on the ROC curve is indicated. The segmented regions are shown superimposed on the images. If a region is abnormal according to the radiologist, the outline of the segmentation is in white instead of black.

Fig. 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.

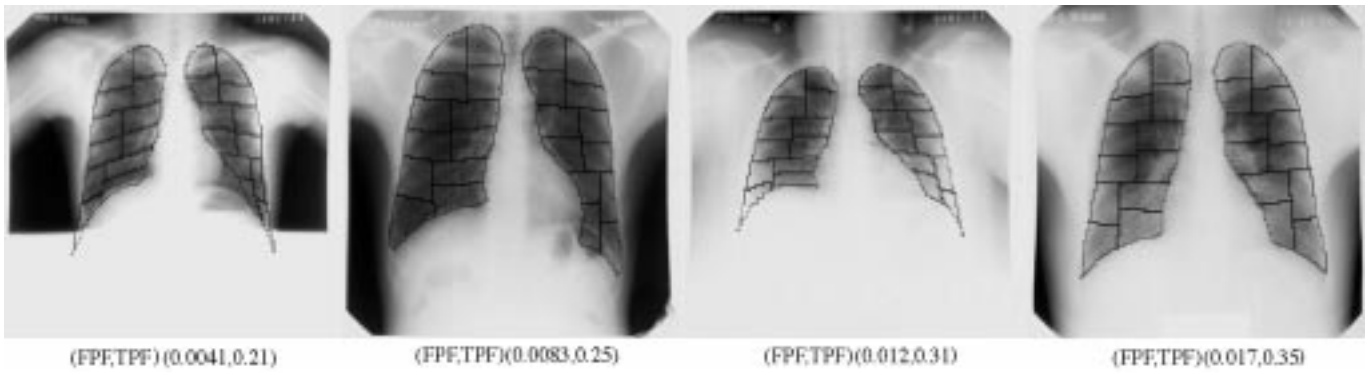


Fig. 6. Of all normal cases in the TB database, the four images displayed here are the most abnormal according to the computerized scheme. These are the most difficult negative cases for the CAD scheme. Below each image, its position on the ROC curve indicated. The segmented regions are shown superimposed on the images.

the regions, as shown in Table IV, that do not 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 us 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.

TABLE I
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 FIG. 2 FOR THE REGION NUMBERING SCHEME

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		

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 ASMs, which are based on the assumption of normal distribution of profiles. Apart from leading to incorrect 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 II. The overall effect of segmentation errors on system performance could be tested by using manual segmentations. Although the segmentation performance of the ASM scheme was adequate for this application, involving regional analyses, it might not be acceptable in applications requiring a high accuracy in terms of lung boundary.

Table III 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 which appear to contain background structures, such as rib borders and large vessels. However, if these areas cannot be accurately detected, and/or abnormal structures are mistaken for normal background structures, the performance may actually deteriorate. Alternatively, preprocessing techniques to remove background structures could be employed, such as rib filtering [30] or contra-lateral subtraction [29].

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.

Note that the total number of features is not much smaller than the number of data sets, especially for the ID database (150 features for 388 cases in the TB database and 186 for 200 cases in the ID database). This may lead to overfitting and the presented results should be interpreted with care. However, the performance of small feature sets is not very different from sets with many features. Several sets of only six features for the ID database already outperform every feature set for the TB database, so the better performance of the method on the ID database cannot be attributed to the larger number of features used.

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 [52] 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 I). 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

TABLE II
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

	TB datase				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 III
TOTAL CLASSIFICATION RESULTS, GIVEN AS A_z, AREA UNDER THE ROC CURVE, FOR VARIOUS COMBINATIONS OF FEATURES FOR THE TB DATABASE AND THE ID DATABASE. FEATURE SETS ARE COMPOSED OF DENSITY FEATURES AND VARIOUS MOMENTS m FROM RESPONSES FROM ALL SIX FILTERS FOR VARIOUS SCALES σ (SEE TEXT FOR DETAILS). 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

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
		$\times$	0.589 ± 0.030	0.633 ± 0.029	0.629 ± 0.030
2,4	2	$\times$	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
		$\times$	0.745 ± 0.034	0.695 ± 0.036	0.787 ± 0.032
1,2,4,8	2,3,4	$\times$	0.983 ± 0.0071	0.847 ± 0.027	0.970 ± 0.011
1,2,4,8,16	2,3,4	$\times$	0.983 ± 0.0072	0.856 ± 0.026	0.968 ± 0.012

TABLE IV
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 SECOND, THIRD, AND FOURTH MOMENTS WITH $\sigma = 1, 2, 4$. PERFORMANCE IS GIVEN AS A_z, AREA UNDER THE ROC CURVE. SEE FIG. 2 FOR THE REGION NUMBERING SCHEME

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		

about 25.000 screenings. This emphasizes the need to establish huge databases in order to obtain progress in CAD.

Our results on the ID database are comparable to, or even slightly better than, those reported by Ishida [18] and Katsuragawa [20] and co-workers. Differences in A_z are in the order of 0.01. 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 finding 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 abnormals 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. In a more advanced system for automating TB screening, the presented method would be only one of the modules. Apart from automatic detection, the system could be used as a second opinion to assist radiologists in reading chest radiographs with textural abnormalities. Whether the system would improve radiologist performance could be tested in observer studies.

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 Fig. 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 Fig. 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 $\times$ 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 Fig. 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.

VI. 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 seg-

mented automatically. Features are extracted from histograms of the responses of a multiscale filter bank. Each region is processed independently with a different k-NN 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.

ACKNOWLEDGMENT

The authors would like to thank the members of the Kurt Rossmann Laboratories, especially Q. Li and R. Engelmann, for help and discussions. Radiographs were annotated by the radiologists W. Haddad and M. Brouwer of the GGD Tilburg, The Netherlands, and H. Abe of the K. Rossmann Laboratories. The authors also thank T. van Walsum of The Image Sciences Institute for his contributions to the imageXplorer environment in which the software for this study has been developed.

REFERENCES

[1] World Health Organization, "The World Health Report 1998: Life in the 21st Century—A Vision for All,", 1998.

[2] G. D. Tourassi, "Journey toward computer-aided diagnosis: Role of image texture analysis," *Radiology*, vol. 213, pp. 317–320, 1999.

[3] R. N. Sutton and E. L. Hall, "Texture measures for automatic classification of pulmonary disease," *IEEE Trans. Comput.*, vol. COMM-21, pp. 667–676, 1972.

[4] G. Revesz and H. L. Kundel, "Feasibility of classifying disseminated pulmonary diseases based on their Fourier spectra," *Investigat. Radiol.*, vol. 8, pp. 345–349, 1973.

[5] R. P. Kruger, W. B. Thompson, and A. F. Turner, "Computer diagnosis of pneumoconiosis," *IEEE Trans. Syst., Man Cybern.*, vol. SMC–4, p. 40, 1974.

[6] E. L. Hall, W. O. Crawford, Jr., and F. E. Roberts, "Computer classification of pneumoconiosis from radiographs of coal workers," *IEEE Trans. Biomed. Eng.*, vol. BME-22, pp. 518–527, June 1975.

[7] J. R. Jagoe and K. A. Paton, "Reading chest radiographs for pneumoconiosis by computer," *Br. J. Ind. Med.*, vol. 32, pp. 267–272, 1975.

[8] ——, "Measurement of pneumoconiosis by computer," *IEEE Trans. Comput.*, vol. COMM-25, pp. 95–97, 1976.

[9] R. J. Tully, R. W. Connors, C. A. Harlow, and G. S. Lodwick, "Toward computer analysis of pulmonary infiltration," *Investigat. Radiol.*, vol. 13, pp. 298–305, 1978.

[10] J. R. Jagoe, "Gradient pattern coding—An application to the measurement of pneumoconiosis in chest X-rays," *Comput. Biomed. Res.*, vol. 12, pp. 1–15, 1979.

[11] X. Chen, K. Doi, S. Katsuragawa, and H. MacMahon, "Automated selection of regions of interest for quantitative analysis of lung textures in digital chest radiographs," *Med. Phys.*, vol. 20, no. 4, pp. 975–982, 1993.

[12] S. Katsuragawa, K. Doi, and H. MacMahon, "Image feature analysis and computer-aided diagnosis in digital radiography: Detection and characterization of interstitial lung disease in digital chest radiographs," *Med. Phys.*, vol. 15, no. 3, pp. 311–319, 1988.

[13] ——, "Image feature analysis and computer-aided diagnosis in digital radiography: Classification of normal and abnormal lungs with interstitial disease in chest images," *Med. Phys.*, vol. 16, no. 1, pp. 38–44, 1989.

[14] J. Morishita, K. Doi, S. Katsuragawa, L. Monnier-Cholley, and H. MacMahon, "Computer-aided diagnosis for interstitial infiltrates in chest radiographs: Optical-density dependence of texture measures," *Med. Phys.*, vol. 22, no. 9, pp. 1515–1522, 1995.

[15] L. Monnier-Cholley, H. MacMahon, S. Katsuragawa, J. Morishita, T. Ishida, and K. Doi, "Computer-aided diagnosis for detection of interstitial opacities on chest radiographs," *Amer. J. Roentgenol.*, vol. 171, pp. 1651–1656, 1998.

[16] S. Katsuragawa, K. Doi, H. MacMahon, L. Monnier-Cholley, J. Morishita, and T. Ishida, "Quantitative analysis of geometric pattern features of interstitial infiltrates in digital chest radiographs: Preliminary results," *J. Digital Imag.*, vol. 9, no. 3, pp. 137–144, 1996.

[17] S. Kido, J. Ikezoe, H. Naito, M. Masuike, S. Tamura, and T. Kozuka, "An image analyzing system for interstitial lung abnormalities in chest radiography," *Radiographics*, vol. 15, pp. 1457–1464, 1995.

[18] T. Ishida, S. Katsuragawa, K. Ashizawa, H. MacMahon, and K. Doi, "Application of artificial neural networks for quantitative analysis of image data in chest radiographs for detection of interstitial lung disease," *J. Digital Imag.*, vol. 11, no. 4, pp. 182–192, 1998.

[19] S. Kido, J. Ikezoe, H. Naito, S. Tamura, and S. Machi, "Fractal analysis of interstitial lung abnormalities in chest radiography," *Radiographics*, vol. 15, pp. 1457–1464, 1995.

[20] S. Katsuragawa, K. Doi, H. MacMahon, L. Monnier-Cholley, T. Ishida, and T. Kobayashi, "Classification of normal and abnormal lungs with interstitial diseases by rule-based method and artificial neural networks," *J. Digital Imag.*, vol. 10, no. 3, pp. 108–114, 1997.

[21] R. M. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE*, vol. 67, pp. 786–804, May 1979.

[22] L. van Gool, P. Dewaele, and A. Oosterlinck, "Survey: Texture analysis anno 1983," *Computer Vision, Graphics, and Image Processing*, vol. 29, pp. 336–357, 1985.

[23] R. T. Reed and J. M. H. du Buf, "A review of recent texture segmentation and feature extraction techniques," *Comput. Vis. Graphics Image Processing*, vol. 57, no. 3, pp. 359–372, 1993.

[24] J. Malik and P. Perona, "Preattentive texture discrimination with early vision mechanisms," *J. Opt. Soc. Amer.-A*, vol. 7, no. 5, pp. 923–932, 1990.

[25] A. C. Bovik, M. Clark, and W. S. Geisler, "Multichannel texture analysis using localized spatial filters," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 55–73, Jan. 1990.

[26] K. D. Laws, "Texture image segmentation," Ph.D. dissertation, University of Southern California, 1980.

[27] M. Unser, "Local linear transforms for texture measurements," *Signal Processing*, vol. 11, pp. 61–79, 1986.

[28] M. Unser and M. Eden, "Multi-resolution feature extraction and selection for texture segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 717–728, July 1989.

[29] Q. Li, S. Katsuragawa, T. Ishida, H. Yoshida, S. Tsukuda, H. MacMahon, and K. Doi, "Contralateral subtraction: A novel technique for detection of asymmetric abnormalities on digital chest radiographs," *Med. Phys.*, vol. 27, no. 1, pp. 47–55, 2000.

[30] F. Vogelsang, F. Weiler, J. Dahmen, M. W. Kilbinger, B. Wein, and R. W. Günther, "Detection and compensation of rib structures in chest radiographs for diagnose assistance," *Proc. SPIE*, vol. 3338, pp. 774–785, 1998.

[31] S. G. Armato, M. L. Giger, and H. MacMahon, "Automated lung segmentation in digitized postero–anterior chest radiographs," *Academic Rad.*, vol. 4, pp. 245–255, 1998.

[32] X. W. Xu and K. Doi, "Image feature analysis for computer-aided diagnosis: Accurate determination of ribcage boundary in chest radiographs," *Med. Phys.*, vol. 22, no. 5, pp. 617–626, 1995.

[33] ——, "Image feature analysis for computer-aided diagnosis: Detection of right and left hemidiaphragm edges and delineation of lung field in chest radiographs," *Med. Phys.*, vol. 23, no. 9, pp. 1613–1624, 1996.

[34] J. Duryea and J. M. Boone, "A fully automatic algorithm for the segmentation of lung fields in digital chest radiographic images," *Med. Phys.*, vol. 22, no. 2, pp. 183–191, 1995.

[35] E. Pietka, "Lung segmentation in digital chest radiographs," *J. Digital Imag.*, vol. 2, pp. 79–84, 1994.

[36] M. S. Brown, L. S. Wilson, B. D. Doust, R. W. Gill, and C. Sun, "Knowledge-based method for segmentation and analysis of lung boundaries in chest X-ray images," *Computerized Med. Imag. Graphics*, vol. 22, pp. 463–477, 1998.

[37] M. F. McNitt-Gray, H. K. Huang, and J. W. Sayre, "Feature selection in the pattern classification problem of digital chest radiograph segmentation," *IEEE Trans. Med. Imag.*, vol. 14, pp. 537–547, June 1995.

[38] O. Tsujii, M. T. Freedman, and S. K. Mun, "Automated segmentation of anatomic regions in chest radiographs using an adaptive-sized hybrid neural network," *Med. Phys.*, vol. 25, no. 6, pp. 998–1007, 1998.

[39] N. F. Vittitoe, R. Vargas-Voracek, and C. E. Floyd, Jr., "Identification of lung regions in chest radiographs using Markov Random Field modeling," *Med. Phys.*, vol. 25, no. 6, pp. 976–985, 1998.

[40] ——, "Markov random field modeling in posteroanterior chest radiograph segmentation," *Med. Phys.*, vol. 26, no. 8, pp. 1670–1677, 1999.

[41] B. van Ginneken and B. M. ter Haar Romeny, "Automatic segmentation of lung fields in chest radiographs," *Med. Phys.*, vol. 27, no. 10, pp. 2445–2455, 2000.

[42] T. F. Cootes, C. J. Taylor, D. Cooper, and J. Graham, "Active shape models—Their training and application," *Comput. Vis. Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[43] G. Behiels, D. Vandermeulen, F. Maes, P. Suetens, and P. Dewaele, "Active shape model-based segmentation of digital X-ray images," in *Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 1999, vol. 1679, MICCAI '99, pp. 128–137.

[44] A. Kelemen, G. Székely, and G. Gerig, "Elastic model-based segmentation of 3-d neuroradiological data sets," *IEEE Trans. Med. Imag.*, vol. 18, pp. 828–839, Oct. 1999.

[45] T. F. Cootes and C. J. Taylor, "Statistical models of appearance for computer vision," Wolfson Image Analysis Unit, Univ. Manchester, Manchester, U.K., Tech. Rep., 1999.

[46] I. Dryden and K. V. Mardia, *The Statistical Analysis of Shape*. London, U.K.: Wiley, 1998.

[47] D. Ruprecht and H. Müller, "Image warping with scattered data interpolation," *IEEE Comput. Graphics Applicat.*, vol. 15, pp. 37–43, Mar. 1995.

[48] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever, "The Gaussian scale-space paradigm and the multiscale local jet," *Int. J. Comput. Vis.*, vol. 18, no. 1, pp. 61–75, 1996.

[49] S. Arya and D. M. Mount, "Approximate nearest neighbor queries in fixed dimensions," in *Proc. 4th ACM-SIAM Symp. Discrete Algorithms*, 1993, pp. 271–280.

[50] C. E. Metz, "ROC methodology in radiologic imaging," *Investigat. Radiol.*, vol. 21, no. 9, pp. 720–733, 1986.

[51] ——, "Some practical issues of experimental design and data analysis in radiological ROC studies," *Investigat. Radiol.*, vol. 24, no. 3, pp. 234–245, 1989.

[52] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Pattern Anal. Machine Intell.*, vol. 22, pp. 4–37, Jan. 2000.