Available online at www.sciencedirect.com
SCIENCE DIRECT
Medical Image Analysis 10 (2006) 247–258
MEDICAL IMAGE ANALYSIS
www.elsevier.com/locate/media

A computer-aided diagnosis system for detection of lung nodules in chest radiographs with an evaluation on a public database

Arnold M.R. Schilham *, Bram van Ginneken, Marco Loog

Image Sciences Institute, University Medical Center Utrecht, The Netherlands

Received 1 July 2004; received in revised form 21 February 2005; accepted 15 September 2005
Available online 15 November 2005

Abstract

A computer algorithm for nodule detection in chest radiographs is presented. The algorithm consists of four main steps: (i) image preprocessing; (ii) nodule candidate detection; (iii) feature extraction; (iv) candidate classification. Two optional extensions to this scheme are tested: candidate selection and candidate segmentation. The output of step (ii) is a list of circles, which can be transformed into more detailed contours by the extra candidate segmentation step. In addition, the candidate selection step (which is a classification step using a small number of features) can be used to reduce the list of nodule candidates before step (iii).

The algorithm uses multi-scale techniques in several stages of the scheme: Candidates are found by looking for local intensity maxima in Gaussian scale space; nodule boundaries are detected by tracing edge points found at large scales down to pixel scale; some of the features used for classification are taken from a multi-scale Gaussian filterbank. Experiments with this scheme (with and without the segmentation and selection steps) are carried out on a previously characterized, publicly available database, that contains a large number of very subtle nodules. For this database, counting as detections only those nodules that were indicated with a confidence level of 50% or more, radiologists previously detected 70% of the nodules.

For our algorithm, it turns out that the selection step does have an added value for the system, while segmentation does not lead to a clear improvement. With the scheme with the best performance, accepting on average two false positives per image results in the identification of 51% of all nodules. For four false positives, this increases to 67%. This is close to the previously reported 70% detection rate of the radiologists.

Keywords: Computer-aided diagnosis; Multi-scale techniques; Chest radiography; Pulmonary nodules; Lung cancer

1. Introduction

Lung cancer is the second most common cancer among both men and women. In the Cancer Facts and Figures 2003 report (American Cancer Society, 2003), the American Cancer Society estimated that in 2003, lung cancer would account for about 13% of all cancer diagnoses and 28% of all cancer deaths. The combined five-year survival rate of lung cancer for all stages is only 15%. If the disease is detected while it is still localized, this rate increases to 49%. However, only 15% of diagnosed lung cancers are at this early stage.

Although not yet proven, it seems that early detection is the most promising strategy to enhance a patient's chance of survival. Early detection can be achieved in a population screening; the most common screenings for lung cancer make use of chest projection radiography, or low-radiation dose Computer Tomography (CT) scans. It has been shown in the Early Lung Cancer Action Project that low-dose CT is more effective than conventional chest X-ray for the detection of pulmonary nodules (Henschke et al., 1999). However, there remains a large incentive to improve upon the detection of nodules in projection X-ray images of the thorax, because the traditional, low-cost chest study

* Corresponding author.
E-mail address: arnold@isi.uu.nl (A.M.R. Schilham).

1361-8415/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.media.2005.09.003

is still by far the most common type of radiological procedure. Moreover, it was found in a lung cancer screening for heavy smokers, that when radiographs were checked in retrospect, 90% of peripheral lung cancers nodules were visible (Muhm et al., 1983). This means that these cancers could have been diagnosed, giving plausible grounds for litigation. In fact lung cancer missed on chest radiographs is the second most common reason for litigation against radiologists in the United States (White and Meyer, 1998). For these reasons there is a particular interest for the development of computer algorithms that can serve as a second reader, highlighting suspicious regions in the radiographs that then have to be judged by a radiologist.

At present there is one commercial computer-aided diagnosis (CAD) system for the detection of pulmonary nodules in X-ray images, that has been approved by the Federal Drugs and Food Administration (FDA): RS-2000 (Deus Technologies, Rockville, MD). The fact that FDA approval was obtained suggests that the system has proven its worth in extensive clinical trials. It also suggests that CAD for this specific task is feasible and beneficial.

The difficulties for detecting lung nodules in radiographs are threefold: (i) There is a wide range in nodule sizes: Commonly a nodule diameter can take any value between a few millimeters up to several centimeters. (ii) Nodules exhibit a large variation in density – and hence visibility on a radiograph – (some nodules are only slightly denser than the surrounding lung tissue, while the densest ones are calcified). (iii) Since nodules can appear anywhere in the lung field, they can be obscured by ribs, the mediastinum and structures below the diaphragm, resulting in a large variation of contrast to the background.

In this paper, we present a novel approach for a computer-aided diagnosis (CAD) scheme for detecting pulmonary nodules in chest X-rays. The key ingredient for our scheme is the recognition of the fact that nodule detection is intrinsically a multi-scale problem. We exploit that observation by using Gaussian scale-space techniques (for candidate detection, for candidate segmentation, and for generation of features for classification) to overcome some of the problems stated above. Presently, the aim of this algorithm is to identify the most likely nodule candidates in thorax images to assist the radiologist who diagnoses those images.

2. Related work

Automatic detection of lung cancer has been an active field of research for the last two decades, and has lead to the publication of a wide variety of approaches for nodule detection in radiographs of the chest (see (van Ginneken et al., 2001) for an overview of computer-aided diagnosis in chest radiography). Typically, proposed methods pass through three stages: (i) candidate detection; (ii) feature extraction; (iii) classification. In what follows we give an indication of the mainstream of procedures that have been used for these steps in related articles.

For the first step, most often the difference-image technique originally proposed by Giger et al. (1988, 1990), or a variant thereof (e.g., Carreira et al. (1998), Keserci and Yoshida (2002)), is used. In the difference-image technique the original image is filtered twice: Once with a spherical kernel to obtain a nodule-enhanced image, and once a median filter is applied to obtain a nodule suppressed image. Nodule candidates are then obtained by thresholding the subtraction of the two filtered images.

Commonly, for the classification step, either a rule-based classifier (e.g., Giger et al. (1990), Xu et al. (1997), Carreira et al. (1998), Li et al. (2001)), or an artificial neural network is used (e.g., Keserci and Yoshida (2002)).

However, the largest part of the variability between published models resides in the feature extraction step. Here, the methods use different sets of meaningful characteristics that ought to enable the classifier to distinguish between actual nodules and false positives. Used feature sets are mostly derived from histogram information (Sklansky and Petković, 1984; Giger et al., 1990), filter outputs (e.g., Keserci and Yoshida (2002)), or descriptions of candidate shape (Sankar and Sklansky, 1982; Carreira et al., 1998; Li et al., 2001). Wei et al. (2002) investigated the feasibility of finding an optimal set of features, derived as a subset from sets of these three classes of features.

A comparison of performances of different CAD schemes is only meaningful if they have been tested on similar databases, and if both the sensitivity of the system and the generated number of false positives are provided. We could identify only a few studies that make use of the same image database that we use (or a database that might be comparable), and that do report a CAD performance by sensitivity at an average number of false positives per image. Freedman et al. (2002) describe a study with RS-2000 using an image database for which observers obtained an area A_z under the ROC curve (Swets, 1997) which is similar to what is obtained by the observers of the JSRT database: respectively, $A_z = 0.835$ and $A_z = 0.833$. This suggests that from an observer's point of view the databases are comparable. In that study it is reported that RS-2000 detects 66% of the nodules with on average 5 false positives per image (Freedman et al., 2002). Wei et al. (2002) reported a sensitivity of 80% at 5.4 false positives per image for the JSRT database. Coppini et al. (2003) also used the JSRT database and found a sensitivity of 60% at 4.3 false positives per image. In Section 7 these two systems are compared to our system.

The novelty of our system is that we regard the problem of nodule detection as an intrinsically multi-scale problem. We exploit that viewpoint by choosing techniques for the various stages of the system that make use of this multi-scale character. For candidate detection we use a multi-scale detector of bright spots (see Section 4.2), and most of the features we use are taken from a multi-scale Gaussian filterbank (see Section 4.4). Furthermore, in an additional stage of the system, where we try to improve details of the outline of the nodule candidates, we make

use of a multi-scale edge-focussing technique (see Section 4.3). As such, the approach of our CAD system is novel and the choice of its building blocks is consistent.

3. Materials

To facilitate future comparison of the performance of our method to that of others, we used the images from the publicly available JSRT database (Shiraishi et al., 2000). These images are digitized to 12 bits posterior–anterior chest radiographs, scanned at a resolution of 2048×2048 pixels; the size of one pixel is 0.175×0.175 mm^2. The database contains 93 normal cases and 154 X-ray images with a proven lung nodule (100 malignant ones). Diameters and positions of the nodules are provided for. The nodules in this database are representative of the problems described in the introduction: The nodule diameters range from 5 to 60 mm (median = 15 mm), they are located throughout the lungs (also behind the heart and under the diaphragm), and their intensities (densities) vary from nearly invisible to very bright. As can be seen in Table 1, the nodules are subdivided in five categories, based on the degree of subtlety for detection, which is influenced by the nodule size, occlusion by other structures and nodule density. The assigned subtlety was based on the consensus of three chest radiologists. The subtlety classes were characterized by the average areas A_z under the ROC curves (Swets, 1997) for 20 radiologists in an observer study. For the experiments described in this paper, we used all the images in the JSRT database, both the normal and the diseased cases. For simplicity, we redivided the nodules into two classes for the experiments: 'practicable' for obvious, relatively obvious, and subtle cases and 'hard' for very subtle and extremely subtle cases.

Table 1
The nodules of the JSRT database are subdivided in five categories

Category	A_z	Number
Obvious	0.990	12
Relatively obvious	0.960	38
Subtle	0.876	50
Very subtle	0.753	29
Extremely subtle	0.568	25
Practicable	0.922	100
Hard	0.667	54

For each category the average area under the ROC curves (A_z) for the radiologists who characterized the images and the number of cases is given [data taken from Shiraishi et al. (2000)]. For the experiments the nodules are recategorized as either hard or practicable.

4. Methods

Typically, a nodule is a roughly spherical object and has a density comparable to water, which is higher than the surrounding lung parenchyma. Consequently, nodules appear as bright, more or less circular spots in chest radiographs. The nodule detection task can thus be achieved by separating all light blobs (i.e., groups of adjacent pixels with similar attenuation) in the images, followed by a categorization of blobs into nodules and non-nodules. The separation process in turn can be split up into a detection part and a segmentation part; in the segmentation step the circles found by the detector are replaced by more detailed outlines. In general, both detection and segmentation will be preceded by an image preprocessing procedure. The last stages of the categorization step will consist of feature extraction and classification of objects in feature space. In Fig. 1 the complete scheme is shown in a flowchart. In the remainder of this section, detailed information regarding the various elements of this scheme will be given.

4.1. Image preprocessing

The efficacy of the detection and segmentation stages in the CAD scheme is boosted by preprocessing of the input

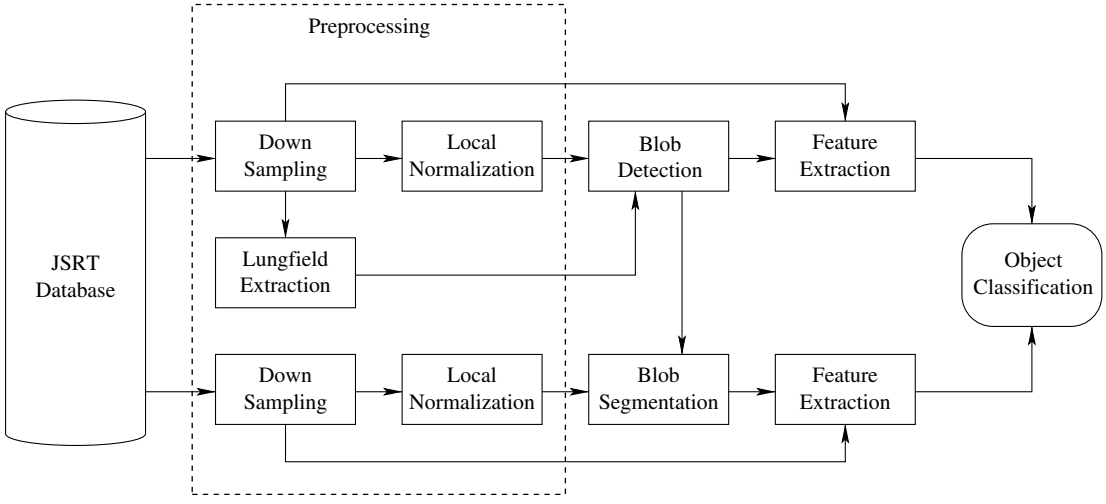

Fig. 1. The multi-scale nodule detection scheme.

images. For both steps, images are down-sampled to reduce the computational effort needed and filtered to obtain comparable contrast throughout each image (see Section 4.1.2). Also, the detector output is confined to a region of interest.

For detection, the images are down-sampled from 2048×2048 to 256×256. The main point of blob segmentation is to find more accurate delineations of the nodule candidates than those produced by the detector. Therefore, the images used for segmentation are taken at a higher resolution: 1024×1024 pixels.

To avoid the turn up of nodule candidates outside the lungs, the output of the blob detector is restricted to the lung fields. Actually the latter procedure was implemented as a postprocessing step for the blob detector, but since the lungfields are determined (see Section 4.1.1) before detection takes place, it is covered in this article as a preprocessing operation.

4.1.1. Lung field segmentation

Lung fields are segmented with an active shape model (ASM) (Cootes et al., 1995). The ASM segmentation scheme requires manually segmented training images, for which we used a set of 230 chest radiographs obtained from a tuberculosis study. The settings of the ASM scheme are those used in (van Ginneken et al., 2002). As usual, the lung fields in chest radiographs are defined as those parts of the lung unobscured by the heart, the mediastinum and the structures below the diaphragm (see Fig. 2 for a segmented lung field for a case of the JSRT database). However, a significant part of the lungs is actually obscured by these structures. If a nodule is located in those parts of the lungs, it will be missed by our detection system. It is possible to extend the region of interest to include those areas, but as there is only a small number of these cases in the JSRT database, we accepted the exclusion of those from the detectable nodules. These images are not excluded from the database, and hence they will end up as nodule specimens missed by the system. This way, we expect that the quoted performance of our method will be closer to what would be found in a real clinical trial, in which such cases will show up.

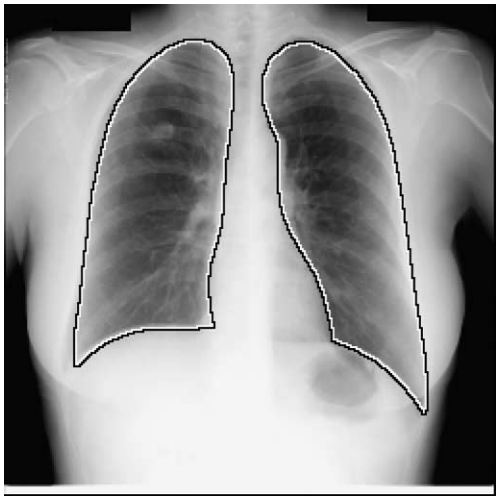

Fig. 2. The lung field of JSRT case JPCLN006 as segmented by an active shape model. The heart, the mediastinum and the diaphragm define the inner boundaries of the lung field.

4.1.2. Local normalization

Ofttimes on a chest X-ray, a nodule has a poor contrast to the background. Not only can a nodule be intrinsically hard to distinguish, being very small, or having a very low density, but frequently the nodule is partly obscured by structures, such as ribs and vessels. By local normalization (LN) filtering, a global equalization of contrast throughout an image is achieved. This filtering also normalizes edge strengths, which enhance the performance of the blob detector (see Section 4.2) and improves the process of segmentation (see Section 4.3).

Locally normalizing an image L constitutes to the following:

$$L_{\mathrm{LN}} = (L - \widetilde{L}) / \left(\widetilde{L^2} - \left(\widetilde{L} \right)^2 \right)^{1/2}, \tag{1}$$

where a tilde indicates Gaussian blurring. Put into words, the local deviation of the image intensity from the local average is normalized on the local standard deviation. The only parameter in this process is the scale of the localization, i.e., the scale parameter σ_{LN} of the Gaussian blur. Fig. 3 shows the result of the local normalization filter applied to an image of the JSRT database.

In the proposed CAD scheme, LN images are used twice: for blob detection and for segmentation. For these steps we used $\sigma_{\mathrm{LN}} = 8$ and 25 pixels, respectively; we took the width of a rib as the σ_{LN} scale, which is typically 8 pixels in the 256×256 images used for detection and amounts to 25 pixels for the 1024×1024 images.

4.2. Multi-scale blob detection

Nodule detection is inherently a multi-scale problem, because the nodules come in many different sizes. For example, the nodules in the JSRT database range in diameter from 6 to 60 mm. In the CAD algorithm Lindeberg's multi-scale detection scheme (Lindeberg, 1998) is employed to deal with this range of sizes. The objective of Lindeberg's detector is to find blobs in scale space, i.e., to find extrema of L_{LN} both spatially and in scale.

Lindeberg's scheme looks for extrema of the Laplacian $\triangle L_{\mathrm{LN}}$. The Laplacian is defined here as $\triangle L_{\mathrm{LN}} \equiv \sigma^2(L_{xx} + L_{yy})$, with L_{xx} and L_{yy} denoting the second-order Gaussian derivatives of L_{LN} at scale σ, with respect to x and y. The factor σ^2 allows for comparison of the Laplacian output at different values of σ.

The scale range for the detector covers $\sigma = 1$–16 pixels in ten exponentially spaced levels (i.e., $\sigma_i = \exp[ic]$, for $i = 0, \ldots, 9$ and $c = \ln(16)/9$). For each detected blob, the position in the image and its corresponding scale of detec-

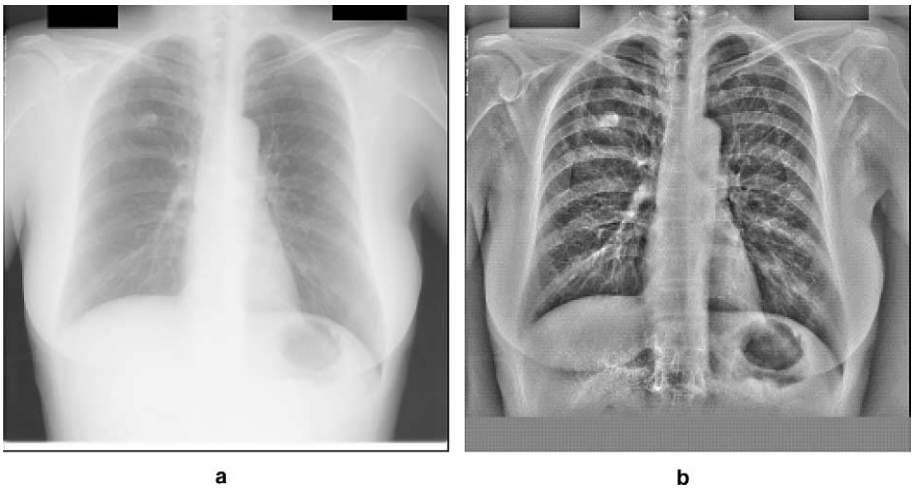

Fig. 3. JSRT case JPCLN006; (a) before and (b) after local normalization. The local normalization scale used was an average width of a rib. The intensities in the lungfields are from 0 to 3824 in (a) and from −2.81 to 2.95 in (b). The bottom part of (b) is set to 0 to illustrate the gray value of 0.

tion σ_D are stored. When blobs overlap only the blob with the highest value of ΔL_{LN} is kept.

For nodule detection we are only interested in bright spots, so the output of this detection step is a list of locations of local maxima with corresponding approximate radii (i.e., detection scales) of the blobs.

A typical example of a nodule as detected by the blob detector is given in Fig. 4. Comparing the detected blob size to the reference standard set by radiologists (e.g., Fig. 4) it appears that the detected blobs are generally too small. Enlarging the detected blob sizes with a constant factor is one way to improve the match between detected sizes and reference sizes; alternatively a blob segmentation step can be added, to find a more detailed outline of the blobs. The latter is discussed in the following section.

4.3. Blob segmentation

A segmentation scheme is deployed that uses the detected blob as a starting point to attempt to better separate the nodule candidate from the surrounding background. As we have no access to the original delineations drawn by the radiologists that formed the basis of the reference standard for the JSRT database, the 'improvement' of nodule segmentation will be evaluated at the end of this section by calculating the overlap between the segmented nodules and the circles formed by their given reference sizes.

The key ingredient of the segmentation scheme is scale-space edge focusing of rays cast through the detected blob centres, akin to the technique used by ter Haar Romeny et al. (1999) for the segmentation of follicles in 3-dimensional ultrasound.

The first step is the construction of Gaussian scale spaces of the data sampled along the cast rays. The extent of the lines is limited to three times the diameter found by the blob detector. A factor three was chosen because we found that detected nodule sizes could be up to three times smaller than the provided reference sizes. Since blobs are brighter than their surroundings, potential blob edges

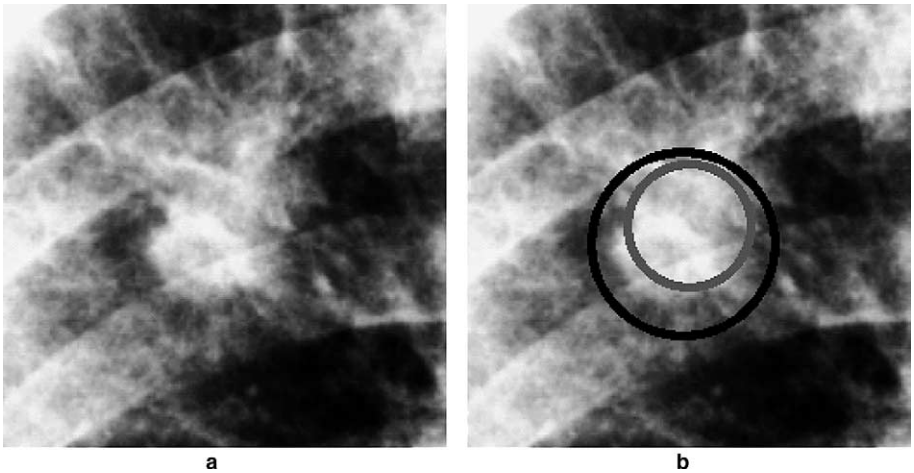

Fig. 4. After histogram equalization, this as 'relatively obvious' categorized nodule (JSRT case JPCLN014) is clear to see (a). In (b), the black circle is representative of the nodule delineation given by the radiologists of the JSRT database, whilst the blob detector finds the grey circle.

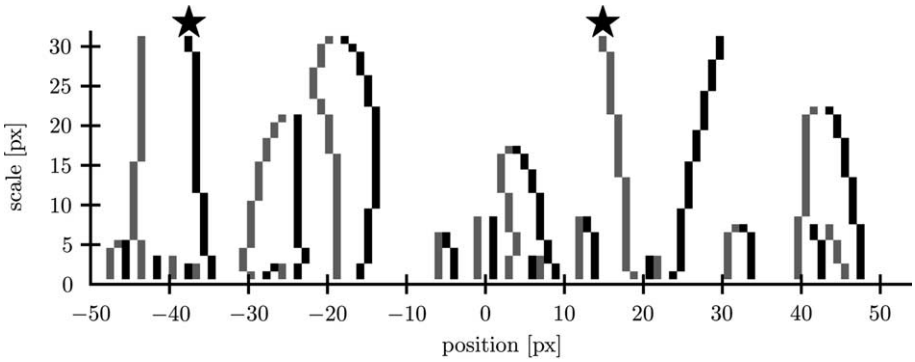

Fig. 5. The scale-space edge focusing technique. The stars indicate the strongest left (black) and right (grey) edges at the maximum scale for edge detection. These two edges are followed down to the smallest scale to find the corresponding locations in the image. Positions (horizontal axis) are relative to the blob centre as detected by the blob detector.

correspond to positive gradients on the left-hand side of the lines ('left' edges) and to negative gradients on the right-hand side of the lines ('right' edges).

Generally, several left and right edges are encountered along a ray at a given scale (see Fig. 5). We choose to pick the strongest edges[1] as identifiers of the boundaries of the blob. These edge points are traced down in scale space to the smallest scale level to find the corresponding edge location in the image. In our application, the scale space spans a scale range from 1 pixel to $1.5 \times \sigma_D$ pixels with increments of 1 pixel. We cast 30 rays in a homogeneous orientation distribution, and trace down a left and a right edge for each orientation, resulting in 60 boundary points per nodule candidate. Notice that this technique can only find convex outlines, so details of nodules that do not have a convex shape will be missed. Since nodules are typically roughly spherical objects, we expect that the error introduced by assuming convex-shaped objects is small.

The rays are cast in images preprocessed with two filters. The first filter is a local normalization with $\sigma_{LN} = 25$ pixels. After LN, it can occur for partly obscured nodules that locally the strongest edges correspond to the boundaries of other structures. This is most likely to happen if the ray travelling outwards from the nodule centre encounters the end of the rib, before it reaches the boundary of the nodule. Because nodules are relatively small, additional structure, they should show up (in LN images) as white, positive blobs, embedded in dark (negative) values. However, the false contour points mentioned above will be found going from a bright region to a bright region with lower intensity values. Thus by applying a filter that increases the edge strength of the zero-crossings in the LN image, the occurrence of these false contour points might be avoided. A simple filter to enhance the edge strength of zero-crossings in the LN images is the following: $L' = -a_1 L_{LN}$ for $L_{LN} < 0$ and $L' = a_2 L_{LN}$ for $L_{LN} \geqslant 0$, with L_{LN} denoting the pixel intensities in the LN image, and a_1 and a_2 positive constants. We find satisfactory results for $a_2/a_1 \geqslant 10$ (typically a_1 is set to 1 and a_2 to 50).

It is unavoidable that some of the found edge points (still) belong to nearby structures rather than the object that is to be segmented. Mostly, these can be identified as distinct outliers when the boundary points of an object are given as a list of distances to the detected blob centre. These outliers can be removed in a postprocessing step. First, each one-dimensional list of distances is median filtered with a kernel half width of 4. Next, the filtered edge points are allowed to grow back to the nearest edge (of the correct sign) if it lies within 10 pixels of the filtered location; if not, the median distance is kept. It is noted that this operation does improve the compactness of the nodule (and hence the overlap with the circular reference standard), but inevitably it does remove some information on the irregularity of the shape of the blob, which might have helped in the distinction between nodules and false positives.

In Fig. 6 the different stages of the blob segmentation process are displayed for a typical nodule, resulting in a compact blob segmentation.

As can be seen in Fig. 7, the overlap of the segmented nodules with the reference circles is significantly improved over the overlap obtained from the detector output. Overlap is defined here as

$$\text{Overlap} \equiv 2\frac{A_S \cap A_R}{A_S + A_R}, \tag{2}$$

with A_S and A_R for the area of the segmented nodule and reference circle, respectively. Also notice that a similar improvement is obtained if one simply multiplies the found σ_Ds with 1.5. In Section 6 the added value of segmentation versus simple scaling of the estimated diameters of the blobs will be assessed.

4.4. Feature extraction

After candidate blobs have been identified, the nodule detection task can be formulated as a classic pattern

[1] Here, edge strength is identified with the magnitude of the gradient at the given scale-space location of the edge.

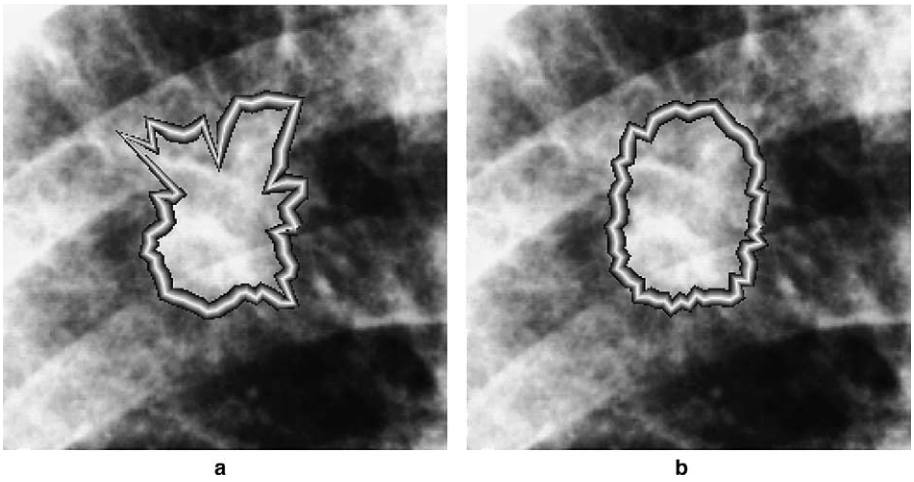

Fig. 6. Segmentation results for the same nodule as shown in Fig. 4. The nodule segmented by the ray casting procedure is shown in (a). After postprocessing the final segmentation of this nodule is given in (b).

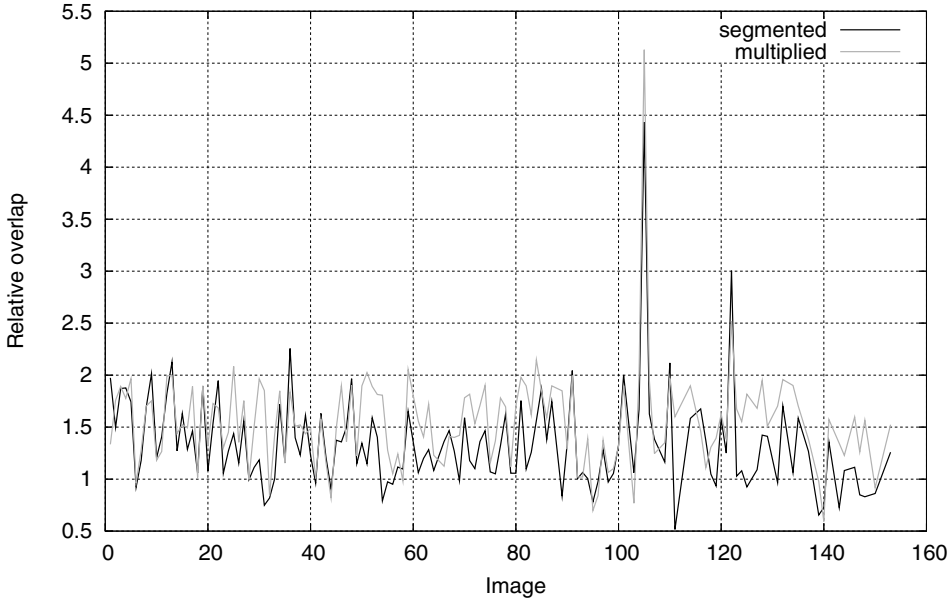

Fig. 7. The overlap between the reference standard delineations and the patches used by the CAD scheme increases if the patches are segmented after detection (black line) and also if the radii of the detected patches are enlarged by a factor 1.5 (grey line). The relative overlap is calculated with respect to the overlap obtained with the detected patches.

recognition task. For classification, the candidates are to be represented by points in a feature space, in which the classifier operates.

For the choice of suitable features to be extracted from the candidate objects, many options have been tried in the literature (see Section 2 for a list of references to various combinations of features). We choose to use features from a multi-scale Gaussian filterbank and a small number of specific features that are readily calculated from the blob detector scheme. The Gaussian filterbank consists of all Gaussian derivatives from 0th to 2nd order for 4 different scales ($\sigma = 1,2,4,8$ pixels). A multi-scale filterbank is chosen to account for the spread in blob sizes. The Gaussian description is used because that is the natural way to calculated regularized derivatives of images. From the filter outputs, the mean and standard deviation are calculated within the segmented area and in a band around the segmented area – defined by doubling all the distances of the boundary points to the detected blob centre. This gives a total of 96 filterbank features. The filterbank is applied to the 1024×1024 images, without local normalization.

Two position features (x_l, y_l) are calculated in a local coordinate system with the centre of mass of the lung fields as the origin and the standard deviations in the x- and y-directions giving the unit lengths along the axes.

From the blob detector the following seven features are taken: The detection scale σ_D and six entities that are derived directly from the Hessian matrix H of L_{LN} (i.e., the 2×2 matrix of all second-order derivatives of the 256×256 locally normalized images multiplied by σ^2.).

These latter six are: L_{xx}, L_{xy}, L_{yy}, $\det H = \lambda_1 \times \lambda_2$, λ_1, and λ_2, with λ_1 and λ_2 denoting, respectively, the largest and smallest eigenvector of H. These 'Hessian' features are evaluated at the detected location and scale of each blob. The combination of four filter features together with σ_D and the Hessian features will be referred to as the 'detector features' in the remainder of this article. These additional four filter features are the mean and standard deviation of the intensity values inside the blob and in a band of width σ_D around the blob.

In summary, the complete set of features consists of 96 filterbank features, 2 position features, and 11 detector features, amounting to total of 109 features. To our knowledge, the usage of a multi-scale Gaussian filterbank to obtain features for classification of nodule candidates has not been reported before.

4.5. Classification

In the final stage of the CAD scheme, for each candidate the probability that it represents an actual nodule is estimated. This probability is given by the posterior probability output of the classifier. The classifier used is a k nearest neighbours (kNN) classifier (Duda et al., 2001), which searches the feature space to find the k nearest neighbours of an object among all nodule candidates from all cases in the database. The posterior probability for being a nodule is then given by n/k where n is the number of actual nodules among the k neighbours. The classification result should not be too strongly dependent on the number of neighbours[2]; a commonly used rule of thumb is $k \simeq \sqrt{\text{number of samples}}$.

For the kNN classifier we used the approximate kNN classifier developed by Arya et al. (1998), with $\epsilon = 2$. The latter means that the distance to the nth nearest neighbour reported by the approximate kNN classifier is less than three times the distance of the true nth neighbour. For k we tried the rule of thumb plus or minus 50% and we saw no clear effect of k on the outcome; therefore for the experiments reported in this paper we used for k the nearest odd integer $\geqslant \sqrt{\text{number of samples}}$.

4.6. Blob selection

Optionally, the classification step can be split-up into a two step process. The extra step is then aimed at removing a large part of the false positives at a low computation cost, before the final classification takes place. For the candidates reduction, a classification is carried out in a 13-dimensional subspace of the 109-dimensional feature space, using the detector features and the positions of the blobs. With this choice of features, the selection step can take place before the segmentation and filterbank feature extraction stages in the algorithm, resulting in a greatly reduced number of computations.

The classifier used for this reduction task is again an approximate kNN classifier ($\epsilon = 2$). The reduction of samples is realized by putting a threshold on the posterior probability to retain on average 20–30 candidates per image. The reduction of candidates resulting from this selection step is demonstrated in Fig. 8 for a case in the JSRT database containing an obvious nodule. For this particular case, selection reduced the number of candidates from 137 to 20.

5. Experiments

In the sketched framework, the simplest scheme would be to run the blob detection, extract all the features from all patches and carry out the classification. This will be called the 'basic scheme' in this article. This scheme can be extended with a blob selection stage and/or a blob segmentation stage, which might enhance the performance of the whole nodule identification task. Table 2 itemizes the four schemes we assess in this paper, and for which the results are compared in Section 6.

In the final classification stages, all 109 features as given in Section 4.4 are taken into consideration. However, the number of features used for classification should be significantly smaller than the number of true training samples, which is 126 after selection (see Table 3 in Section 6). Therefore sequential forward selection (SFS) (Duda et al., 2001) is used to select a subset of a number of features (N_{feat}), with $N_{\text{feat}} \leqslant 20$. The optimization criterion for SFS was sensitivity of the classifier at a fixed specificity level (S_{crit}), i.e., one point of the ROC curve. We picked values for S_{crit} corresponding to optimization of CAD sensitivity at on average 4 false positives per image (one point of the Free Response Receiver Operating Characteristics (FROC curve Bunch et al., 1978)). Explicitly this means that we use $S_{\text{crit}} = (\text{number of images}) \times (\text{average number of FPs per image})/(\text{number of candidates})$.

Analogously, SFS was used in the candidate selection step to take at most 10 from the 13 features for blob selection. In this case we could have used all features, but we want to keep only those features that contribute to a significant reduction of false positives, whilst not degrading the number of true positives too much.

Fivefold cross-validation (Duda et al., 2001) was used for all classification processes: in SFS, for blob selection, and for final classification. Thence the images of the JSRT database were split over five sets, with images of the same subtlety class distributed evenly over the sets. One set is then classified with a classifier trained on the four remaining sets; this is repeated five times to classify all samples.

6. Results

The total number of blobs detected for the whole JSRT database was 32,989, i.e., on average 134 blobs per image.

[2] That is, the ordering for most likely objects for a nodule should be roughly the same if k is varied within sensible limits.

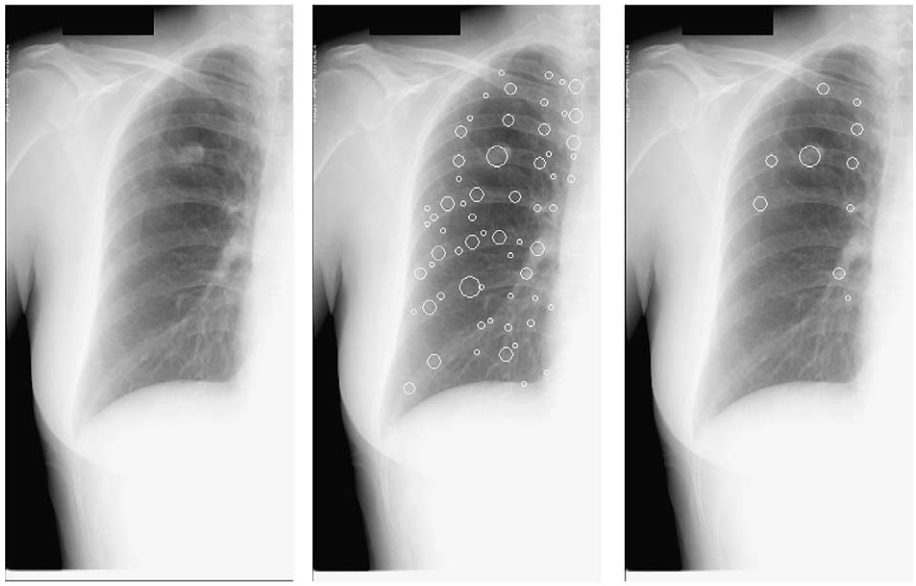

Fig. 8. The detection and selection parts of the multi-scale CAD scheme in action for the right lung of JSRT case JPCLN006. (a) The original image. (b) All the blobs detected by the blob detector superimposed on the original image. (c) The blobs left after blob selection. For this case, the actual nodule (found as the largest blob in (c)) also has the highest probability for being a nodule after the final classification of these blobs both in (b) and in (c).

Table 2
Four variations of the multi-scale detector scheme, differing in the use of a selection step and segmentation stage

Scheme	Selection	Segmentation
Basic	No	No
Segmentation	No	Yes
Selection	Yes	No
Segmentation and selection	Yes	Yes

Table 3
Actual nodules and candidates retained at different stages of the detection scheme

	Number of candidates	Number of nodules
JSRT database	–	154 (100%)
Lung field segmentation	–	141 (91.5%)
Blob detection	32,989	136 (88.3%)
Blob selection	4998	126 (81.8%)

The selection step reduces the number of candidates to 4998. The penalty for this removal of 85% of the candidates is the loss of 10 nodules that were in the list of candidates before selection (i.e., 7% less nodules after selection). Table 3 summarizes the number of candidates and nodules at different stages of the multi-scale detection scheme. A nodule was considered to be detected if there was any overlap between the detected blob and the reference standard of the nodule. This is a rather lenient criterion, but after the detection step the nodule candidate sizes are enlarged (by segmentation or by multiplying σ_D by 1.5) so the blobs used for feature extraction are likely to include a significant part of the corresponding nodules. Note that successful detection is determined in the blob detection step; if a nodule is not detected at this stage, it cannot turn into a successful detection in a later stage.

Note that 13 nodules out of 154 are obstructed by the heart, mediastinum or structures below the diaphragm. Looking at a chest X-ray image (e.g., Fig. 2), it is seen that the projected lung area covered by these obstructing structures is roughly half the area of one lung. The actual lung volume corresponding to this obstructed area is much lower than half of a lung volume; if we assume that the available lung volume is reduced by a factor three, we expect that the probability for a nodule to be in one of the obstructed areas is 1 in 12. Apparently this appearance rate is reflected in the JSRT database. These obscured nodules are outside the analyzed region of the images, and therefore they could not be detected. Five more nodules are missed by the multi-scale blob detection. Four of these missed nodules are 'extremely subtle' or 'very subtle' cases which – if possible at all – are very hard to detect for human observers as well (see Table 1 and Shiraishi et al. (2000)).

In Figs. 9 and 10 the final performances of the four CAD schemes are shown as FROC curves for the JSRT database, measuring sensitivity (overall and for the hard and practicable cases separately) as a function of the average number of false positives per image. Indeed the graphs show that it is much harder to detect nodules of the hard category than those of the practicable class, which reflects the experience of the radiologists.

The sensitivities of the CAD schemes at on average two and four false positives (FPs) per image are listed in Table 4. These two values seem reasonable for a CAD system for nodule detection in radiographs; a system generating too many false positives would probably not be used by a

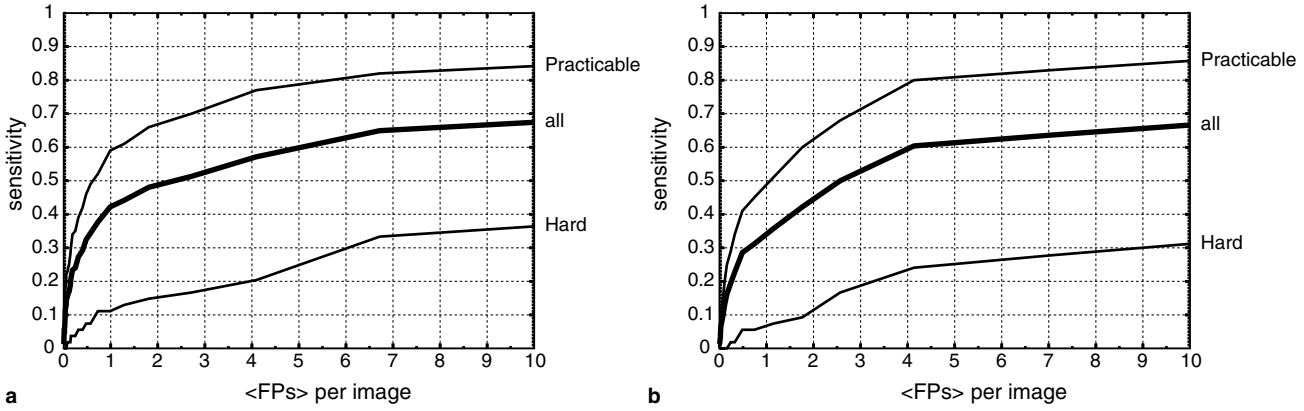

Fig. 9. FROC curves of the system for the complete JSRT database, showing the sensitivity for all nodules, for the practicable nodules, and for the hard nodules; (a) for the basic scheme and (b) for the scheme with segmentation.

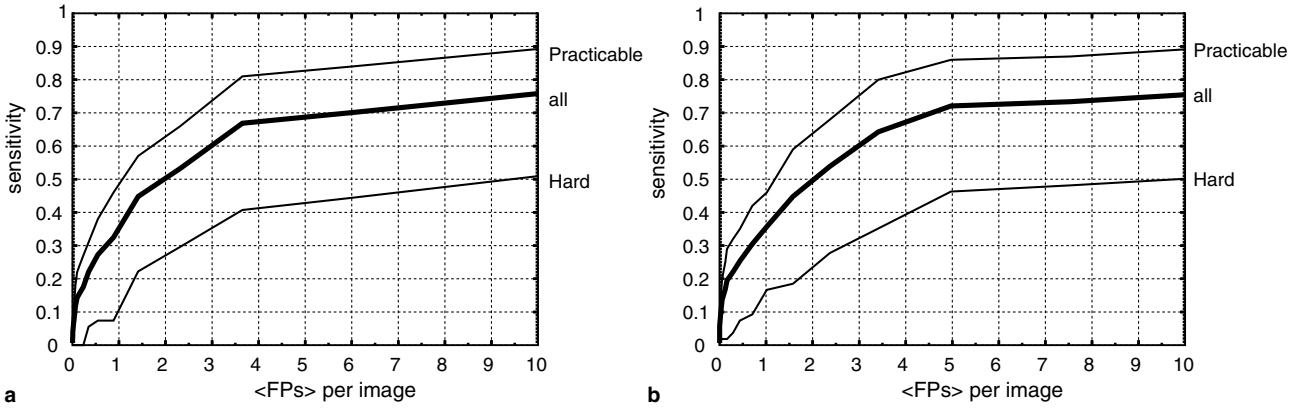

Fig. 10. FROC curves of the system for the complete JSRT database, showing the sensitivity for all nodules, for the practicable nodules, and for the hard nodules; (a) for the scheme with selection and (b) for the scheme with selection and segmentation.

Table 4
Performances of the four CAD schemes expressed as sensitivity when accepting on average two and four false positives per image

Scheme	$\langle FPs \rangle = 2$			$\langle FPs \rangle = 4$		
	All	Practicable	Hard	All	Practicable	Hard
Basic	0.49	0.67	0.15	0.57	0.78	0.20
Segmentation	0.45	0.63	0.12	0.60	0.79	0.24
Selection	0.51	0.63	0.27	0.67	0.82	0.41
Segmentation and selection	0.50	0.64	0.24	0.67	0.82	0.40

radiologist, because it then significantly increases his workload, while the probability that the system is wrong on a per marker basis is very large. The numbers in Table 4 show that the addition of the first selection step is always advantageous. This enhancement is most apparent for the hard cases. The addition of the segmentation stage does not have a clear positive effect on the sensitivity of the scheme at these two operating points. These results prompt us to choose the selection scheme as our CAD system of preference.

7. Conclusions and discussion

We have shown promising results obtained with a multi-scale CAD scheme for the detection of pulmonary nodules in chest radiographs. We have tested the effects of including or excluding two stages of the system: an additional candidate selection step and a candidate segmentation stage. The inclusion of the candidate selection step had a clear positive effect on system performance. The effect of the candidate segmentation step was less apparent. Therefore we recommend to use the system with candidate selection, but to exclude the segmentation stage.

For the evaluation of our algorithm we used the publicly available JSRT database of radiographs Shiraishi et al. (2000), which facilitates the use of our system as a benchmark for future studies. The JSRT database exhibits a realistic distribution of nodules sizes and locations: Diameters ranges from 6 to 60 mm and locations are all over the lungs

(including a realistic fraction of nodules obstructed by the mediastinum and the diaphragm). The observer studies in (Shiraishi et al., 2000) indicated that radiologists find it particularly difficult to detect the very and extremely subtle nodules in the JSRT database; Counting only those detections that observers rated with a confidence of at least 50%, the radiologists detected on average 85% of the practicable cases and 44% of the hard cases (Shiraishi et al., 2000). With on average four false positives per image, our scheme correctly marks 41% of the hard cases and 82% of the practicable cases. This is a very encouraging result that suggests that our method could provide a useful clinical tool. However, this remains to be proven in observer studies.

We have attempted to implement other CAD schemes for comparison. In theory published methods should allow for an independent implementation. In practice this turned out to be impossible; either the training data (or other crucial, supplemental data) is unavailable, or the description of the method is incomplete. The only algorithm that could be implemented was the system by Carreira et al. (1998). Their algorithm consist of two steps: detection of candidates and classification of candidates. After implementing the detection part and running it against the JSRT database, the result was a total of 1557 candidates containing only 54 nodules (i.e., a sensitivity of 0.35 with on average 6 FPs per image); the remaining 100 nodules were not detected at all. Clearly our CAD scheme performs better for the JSRT database.

Comparison of the performance of our CAD scheme to published results of others is only meaningful if the other methods use a similar database and report the same measure for performance. We could find only three reported studies that met these requirements. The RS-2000 system has been tested on a database of similar difficulty as the JSRT database (Freedman et al., 2002) (observers obtained an area A_z score of 0.833 for the JSRT database Shiraishi et al. (2000) and 0.835 for the database in Freedman et al. (2002)). In that particular study RS-2000 detects 66% of the nodules with on average 5 false positives per image. At that operating point, the FROC curves in Fig. 10 show performance rates of 69% and 72%, which would seem to favour our system. However, RS-2000 has proven its worth in over 10,000 cases to obtain FDA approval, whereas our system has only been tested on the JSRT database with only 154 nodule cases. Wei et al. (2002) reported a sensitivity of 80% at 5.4 false positives per image for the JSRT database. However, to reach that performance Wei et al. (2002) had to use 202 uncorrelated features; that is cause for some concern because it means that the system uses more features than the available number of true positive samples in the database, and as a consequence the risk of overtraining the system is high. At 5.4 FPs per image, our CAD scheme reaches 73% sensitivity, and uses less than 20 features. Finally, Coppini et al. (2003) also used the JSRT database and found a sensitivity of only 60% at 4.3 false positives per image for their CAD system. We conclude that from the comparisons that we could make, our system shows the best performance (and compared to Wei et al. (2002) we have at least a safer number of features).

In the proposed methodology we have tried to use natural and efficient operations for all processes in the chain. However, substitution of these operations by other processes that are also specifically developed for the task of nodule detection could result in the same (or better) performance. For example, iris filtering (Keserci and Yoshida, 2002) might have benefits over the use of local normalization, because it is specifically designed to enhance the contrast of compact blobs. Also, the addition of more specialized features is likely to lead to improved classification. What those features might be remains an issue for future research. Presently we have demonstrated a consistent CAD system that uses Gaussian scale-space techniques to capture the multi-scale character of nodules in chest radiographs; these techniques have not been exploited for this task before.

In future research projects we will focus on the reduction of false positives to boost the performance of the system. Careful investigation of the candidates that were falsely classified with a high probability for a nodule, shows that overlapping bony structures (notably ribs with other ribs or scapula) and the hilum generate the most probable false positives (see Fig. 11 for examples). Using not only one classifier for nodules, but using additional classifiers for

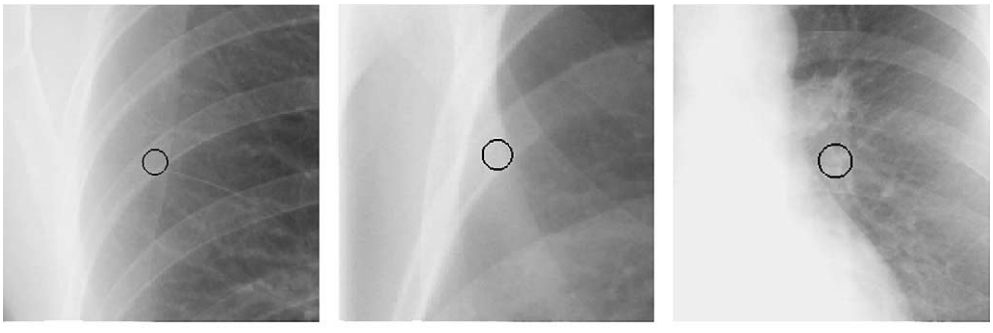

Fig. 11. Examples of false positives of the nodule detection scheme, that had high probabilities for being nodules: overlying bony structures (a) and (b); a part of the hilum (c).

the overlapping bony structures and the hilum could be a good strategy to reduce the number of false positives (see, e.g., Lin et al. (1996)). In addition, since one selection step between detection of candidates and final classification of candidates leads to improved performance, the construction of several selection steps that make classifications with limited sets of features to progressively reduce the number of false positives, may lead to better performance.

Apart from improving the performance of our CAD system, the important question whether our CAD system really improves the sensitivity rates of radiologists should be addressed. Such a validation requires a carefully executed observer study with several radiologists and could not be included in the present article. We want to do this evaluation in a future research.

As a final note, we surmise that the detection framework presented might also prove useful in determining whether a given nodule is benign or malignant (Aoyama et al., 2002). Although the segmentation of candidates did not yield improved results for the detection, it could turn out to be important for characterization. This is definitely something that we will investigate in the near future.

References

American Cancer Society, 2003. Cancer facts and figures 2003. Technical reports, American Cancer Society, Inc., Atlanta, GA.

Aoyama, M., Li, Q., Katsuragawa, S., MacMahon, H., Doi, K., 2002. Automated computerized scheme for distinction between benign and malignant solitary pulmonary nodules on chest images. Medical Physics 29, 701–708.

Arya, S., Mount, D., Netanyahu, N., Silverman, R., Wu, A., 1998. An optimal algorithm for approximate nearest neighbor searching in fixed dimensions. Journal of the ACM 45 (6), 891–923.

Bunch, P.C., Hamilton, J.F., Sanderson, G.K., Simmons, A.H., 1978. A free response approach to the measurement and characterization of radiographic-observer performance. Journal of Applied Photographic Engineering 4, 166–171.

Carreira, M.J., Cabello, D., Penedo, M.G., Mosquera, A., 1998. Computer-aided diagnoses: automatic detection of lung nodules. Medical Physics 25 (10), 1998–2006.

Cootes, T.F., Taylor, C.J., Cooper, D., Graham, J., 1995. Active shape models – their training and application. Computer Vision and Image Understanding 61 (1), 38–59.

Coppini, G., Diciotti, S., Falchini, M., Villari, N., Valli, G., 2003. Neural networks for computer-aided diagnosis: Detection of lung nodules in chest radiograms. IEEE Transactions on Information Technology in Biomedicine 7, 344–357.

Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern Classification, second ed. John Wiley & Sons, New York.

Freedman, M.T., Lo, S.-C.B., Osicka, T., Lure, F., Xu, X.-W., Lin, J., Zhao, H., Zhang, R., 2002. Computer aided detection of lung cancer on chest radiographs: Effect of machine cad false positive locations on radiologists' behavior. In: Proceedings of the SPIE, vol. 4684, pp. 1311–1319.

Giger, M.L., Doi, K., MacMahon, H., 1988. Image feature analysis and computer-aided diagnosis in digital radiography: automated detection of nodules in peripheral lung fields. Medical Physics 15 (2), 158–166.

Giger, M.L., Doi, K., MacMahon, H., Metz, C.E., 1990. Computerized detection of pulmonary nodules in digital chest images: use of morphological filters in reducing false positive detections. Medical Physics 17 (5), 861–865.

Henschke, C.I., McCauley, D.I., Yankelevitz, D.F., Naidich, D.P., McGuinness, G., Miettinen, O.S., Libby, D.M., Pasmantier, M.W., Koizumi, J., Altorki, N.K., Smith, J.P., 1999. Early lung cancer action project: overall design and findings from baseline screening. Lancet 354, 99–105.

Keserci, B., Yoshida, H., 2002. Computerized detection of pulmonary nodules in chest radiographs based on morphological features and wavelet snake model. Medical Image Analysis 6, 431–447.

Li, Q., Katsuragawa, S., Doi, K., 2001. Computer-aided diagnostic scheme for lung nodule detection in digital chest radiographs by use of a multiple-template matching technique. Medical Physics 28, 2070–2076.

Lin, J.-S., Lo, S.C., Hasegawa, A., Freedman, M.T., Mun, S.K., 1996. Reduction of false positives in lung nodule detection using a two-level neural classification. IEEE Transactions on Medical Imaging 15 (2), 206–217.

Lindeberg, T., 1998. Feature detection with automatic scale selection. International Journal of Computer Vision 30 (2), 79–116.

Muhm, J.R., Miller, W.E., Fontana, R.S., Sanderson, D.R., Uhlenhopp, M.A., 1983. Lung cancer detected during a screening program using four-month chest radiographs. Radiology 148, 609–615.

Sankar, P.V., Sklansky, J., 1982. A Gestalt guided heuristic boundary follower for X-ray images of lung nodules. IEEE Transactions on Pattern Analysis and Machine Intelligence 4 (3), 326–331.

Shiraishi, J., Katsuragawa, S., Ikezoe, J., Matsumoto, T., Kobayashi, T., Komatsu, K., Matsui, M., Fujita, H., Kodera, Y., Doi, K., 2000. Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. American Journal of Roentgenology 174, 71–74.

Sklansky, J., Petković, D., 1984. Two-resolution detection of lung tumors in chest radiographs. In: Rosenfeld, A. (Ed.), Multiresolution Image Processing and Analysis. Springer, Berlin, pp. 365–378.

Swets, J., 1997. Roc analysis applied to the evaluation of medical imaging techniques. Investigative Radiology 14, 109–121.

ter Haar Romeny, B.M., Titulaer, B., Kalitzin, S., Scheffer, G., Broekmans, F., te Velde, E., Staal, J.J., 1999. Computer assisted human follicle analysis for fertility prospects with 3D ultrasound. In: IPMI '99Lecture Notes in Computer Science, vol. 1613. Springer-Verlag, Heidelberg, pp. 56–69.

van Ginneken, B., Katsuragawa, S., ter Haar Romeny, B.M., Doi, K., Viergever, M.A., 2002. Automatic detection of abnormalities in chest radiographs using local texture analysis. IEEE Transactions on Medical Imaging 21 (2), 139–149.

van Ginneken, B., ter Haar Romeny, B.M., Viergever, M.A., 2001. Computer-aided diagnosis in chest radiography: A survey. IEEE Transactions on Medical Imaging 20, 1228–1241.

Wei, J., Hagihara, Y., Shimizu, A., Kobatake, H., 2002. Optimal image feature set for detecting lung nodules on chest X-ray images. In: Computer Assisted Radiology and Surgery (CARS 2002). Springer, Berlin, pp. 706–711.

White, C.S., Meyer, C.A., 1998. Missed lung cancer: medicolegal implications. Applied Radiology 27 (8).

Xu, X.-W., Doi, K., Kobayashi, T., MacMahon, H., Giger, M.L., 1997. Development of an improved cad scheme for automated detection of lung nodules in digital chest images. Medical Physics 24, 1395–1403.