Automated Detection of Pulmonary Nodules from Low-Dose Computed Tomography Scans using a Two-Stage Classification System based on Local Image Features

K. Murphy[a], A.Schilham[a], H. Gietema[b], M. Prokop[b], B. van Ginneken[a]

[a]Image Sciences Institute, University Medical Centre Utrecht, The Netherlands;
[b]Department of Radiology, University Medical Centre Utrecht, The Netherlands;

ABSTRACT

The automated detection of lung nodules in CT scans is an important problem in computer-aided diagnosis. In this paper an approach to nodule candidate detection is presented which utilises the local image features of shape index and curvedness. False-positive candidates are removed by means of a two-step approach using kNN classification. The kNN classifiers are trained using features of the image intensity gradients and grey-values in addition to further measures of shape index and curvedness profiles in the candidate regions. The training set consisted of data from 698 scans while the independent test set comprised a further 142 images. At 84% sensitivity an average of 8.2 false-positive detections per scan were observed.

Keywords: Nodule Detection, Low-dose CT, Shape Index, Curvedness, kNN Classification

1. INTRODUCTION

Lung cancer is the leading cause of cancer deaths in the United States, killing more people than the next three most deadly cancers combined (colon cancer, breast cancer and pancreatic cancer).[1] In 2006 the American Cancer Society estimates that 29% of all cancer deaths in the United States will be due to lung cancer. One of the key issues in addressing this statistic is the fact that lung cancer is rarely diagnosed in the early more treatable stage of the disease which is relatively asymptomatic.[1] For this reason routine screening programmes are currently being considered as a possible means to detect the first signs of the disease in apparently healthy subjects.

Early stage lung cancer manifests itself in the form of pulmonary nodules which are visible on chest scans. These nodules are generally approximately spherical in shape, with those which are attached to the pulmonary wall (pleural nodules) being roughly hemispherical. Pulmonary nodules are most easily detected on CT images[2] where they appear as small bright spots, with grey-values very similar to those of blood vessels in the lungs. Although not all pulmonary nodules are malignant or even require treatment, detection of such nodules is the first crucial step in identifying patients who may need follow up treatment or examinations. The examination of a chest CT scan is however a time-consuming and error-prone task for a radiologist. Studies have shown that radiologists frequently fail to detect all visible nodules in a scan.[3, 4] These issues have motivated the development of computer-aided detection (CAD) solutions for lung nodule detection on CT scans. The CAD solution is generally not intended as a replacement for a radiologist, but rather as a very efficient second reader[5] which can quickly identify suspicious structures for his attention.

The area of computer-aided nodule detection in CT is an active field of research with a wide range of approaches having been published in the literature of the last number of years.[6–14] It has proved difficult in the approaches documented to date to reliably detect all nodule types without simultaneously introducing an unacceptable level of false-positives. Since pleural nodules generally have slightly different characteristics than their non-pleural counterparts, a number of authors have found it necessary to employ separate techniques for detecting them.[11, 15, 16] Nodules of an atypical appearance have proved particularly difficult to detect by conventional methods. Such nodules may be non-solid or part-solid, of ground-glass or mixed-ground-glass opacity, spiculated, or of an otherwise irregular appearance. Nodules located in diseased lung tissue which disguises the nodule structure also present problems for automated detection systems.[13]

Although the American Food and Drug Administration (FDA) has approved a commercial nodule detection system for CT scans, the results from the system developed by R2 Technology Inc. are imperfect and modifications to it are ongoing.[17] A study by Lee et al.[6] in 2005 verified that the system requires improvement and that it identified just 60% of 78 pulmonary nodules presented in their tests, with an average of 1.56 false-positive detections per scan. The nodules in question ranged from 4mm-15.4mm in diameter and the study specifically excluded scans exhibiting lung disease, masses larger than 2cm, multiple nodules (more than six), large pleural effusions, non-solid or partly-solid nodules.

In this work a method is presented for detecting nodule candidates in CT scans based on shape index (SI) and curvedness (CV) values. The detected structures are further classified as nodules or otherwise by a series of two kNN-classifiers[18] trained using features of intensity gradient magnitude and orientation, grey value intensity and SI and CV values around the candidate location. The use of local image features in nodule detection results in a system which can identify nodules based on partial regions of their surface without requiring the entire structure to conform to a specific shape or appearance. The features used in kNN-classification are extremely efficient in distinguishing true nodules from false-positives as demonstrated by the sensitivity and specificity rates of the classifiers. In the first classification we achieve 94% sensitivity with 88% specificity, while both sensitivity and specificity reach 90% for the second classifier.

2. MATERIALS

The Nelson Study is an experimental lung cancer screening programme currently taking place in the Netherlands and Belgium. As a participant in this programme the University of Utrecht has access to a large database of low-dose chest CT scans. Each of these scans has been examined by a radiologist and the locations of the observed pulmonary nodules have been annotated. All visible nodules are annotated for the Nelson study without any restrictions on the size of the structure or the necessity for follow up procedures. The data therefore includes extremely small or trivial detections as well as larger and more significant structures. For this work a total of 698 scans were used in the generation of training data while testing was performed on a further 142 scans. The slice size was 512×512 pixels, with the number of slices ranging from 383 to 545. Slice thickness was fixed at 1mm with slice-spacing of 0.7mm. Pixel spacing in the X and Y directions varied from 0.6mm to 0.9mm.

In all cases CT scanning was performed on a 16 detector-row scanner (Mx8000 IDT or Brilliance 16P, Philips Medical Systems, Cleveland, OH, USA). The scans were realized within 12 seconds, in spiral mode with 16×0.75mm collimation, and without contrast-injection. Exposure settings were 30mAs at 120kVp for subjects weighing below 80kg or 30mAs at 140kVp for those weighing over 80kg. Scans were performed in inspiration after appropriate instruction.

3. METHODS

3.1. Overview

Our scheme for nodule detection utilises the shape index (SI) and curvedness (CV) features to detect initial nodule candidates. These are 3D local image features which are calculated per voxel and give insight into the surface topology at every point in the image volume. Figure 1 shows a flowchart of the procedure described below. The individual steps of the scheme will be explained in more detail in the remainder of section 3.

The algorithm begins with some image pre-processing which involves (i) down-sampling of the image to reduce computation time for the nodule detection process and (ii) lung volume segmentation so that the search space for nodule candidates is reduced. The SI and CV values are then calculated for voxels in the lung volume. Thresholding on the SI and CV values is carried out to define seed points whose shape index and curvedness both fall within pre-defined narrow ranges. Next each of the seed points is grown by hysteresis thresholding[19] to form clusters of connected voxels. A cluster at this stage represents a surface, or partial surface in the image volume which has SI and CV values similar to those found on nodule surfaces. The location of such a cluster is defined by the centre of mass of its component voxels. Clusters whose volume is below a defined threshold t_{vol} are discarded.

Clusters in close proximity to each other are now merged since it is highly likely that close neighbouring clusters represent portions of the surface of a single structure. The merging procedure involves the amalgamation of all voxels concerned to form a single cluster and re-calculation of the centre of mass. The resultant merged

structures represent the initial nodule candidates and subsequent steps are aimed at reducing the number of false-positive detections.

For the first false-positive reduction step 8 relatively simple features relating to the cluster geometry and grey-values in the region of the candidates are calculated. A kNN-classifier trained on an independent dataset is used to make an initial classification and remove the most obvious false-positive candidates. The remaining candidate structures are analysed in more detail with 22 features relating to gradient orientation and magnitude, grey value intensities, SI and CV values and geometric cluster properties being calculated for each one. A second kNN-classifier, independently trained, is employed to distinguish likely true nodules from false-positives based on this feature set.

The method of detection is exactly the same for both pleural and non-pleural nodules with two minor exceptions relating to threshold values in each case. These are described in sections 3.4.1 and 3.4.2.
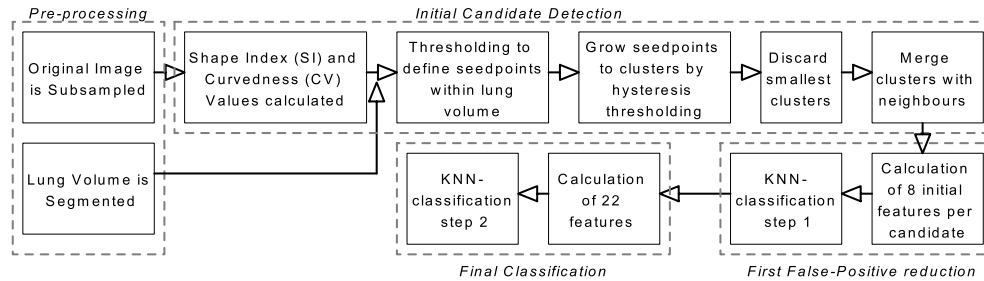
Figure 1. Overview of Nodule Detection Scheme

3.2. Pre-processing

Before beginning with nodule detection some initial processing is carried out on the raw image data. The first step is to down-sample the data to improve the algorithm efficiency. The matrix size of 512×512 in the original Nelson study images is reduced to 256×256, with the number of slices being reduced accordingly.

The second pre-processing step involves the segmentation of the lung volume from the surrounding tissues. The mask obtained from this segmentation is used to ensure that nodule detection is performed within the lung volume only. This process has the two-fold advantage of reducing computation time and preventing the possible detection of false positive structures in regions of the image outside the lungs. Lung segmentation was carried out using an algorithm based on that of Hu et al.[20]

3.3. Shape Index and Curvedness

The SI and CV values at a point are derived from the principal curvatures k_1 and k_2, which are the minimum and maximum curvatures at that point. In fact the feature space spanned by SI and CV is a polar representation of the space spanned by k_1 and k_2.[21] In the context of nodule detection however the SI/CV feature space has the benefit that the topological shape and magnitude of curvature are decoupled allowing them to be treated independently.

Shape Index Shape index (SI) is a measure of topology which relates well to the intuitive understanding of shape. Every distinct shape (with the exception of the plane) corresponds to a unique SI value.[21] These values vary between -1 and 1, where 1 is the value found for a voxel on a perfectly spherical surface and 0.5 the value for a voxel from a cylindrical surface. Figure 2 shows the array of shapes described by varying SI values. In fact, this figure depicts the mapping of various shapes onto the unit circle in the plane defined by k_1 and k_2, with shapes on the unit circle having identical curvedness values. At the centre of the circle where $k_1 = k_2 = 0$ the defined shape is a plane and, as noted above, has no distinct SI value.

A majority of pulmonary nodules can generally be described as having a roughly spherical shape, although naturally real nodules deviate from this model to a lesser or greater degree. The algorithm used in this work enables the detection of structures which are spherical, close to spherical, or even structures which have only some portions of their surface approximating a spherical shape. Since the SI value at a point on

a cylindrical surface such as a vessel is quite different to that on a spherical surface, nodules and vessels are easily distinguished in most cases.
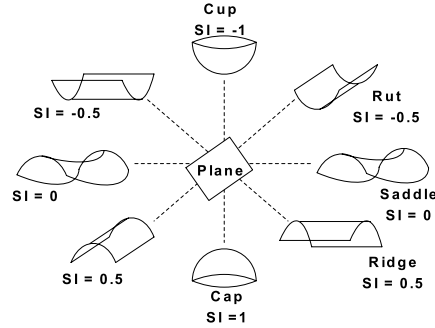

Figure 2. Shapes described by gradually varying Shape Index values

Curvedness The curvedness values in an image vary between 0 and ∞. Due to the limited size range of nodules the CV values on their surface can generally be expected to fall within a particular range. By selecting only voxels whose CV values fall within this range, regions whose shape index values may be of interest but which represent much larger structures are excluded - for example parts of the pleural wall or of the mediastinum.

3.3.1. Calculating Shape Index and Curvedness

The shape index and curvedness values at a voxel are calculated using the principal curvatures k_1 and k_2 at that point as follows.[21] *

$$SI = \frac{2}{\pi}\arctan\left(\frac{k_1 + k_2}{k_1 - k_2}\right) \tag{1}$$

$$CV = \sqrt{k_1{}^2 + k_2{}^2} \tag{2}$$

The principal curvatures k_1 and k_2 are calculated using first and second order derivatives of the image blurred with a Gaussian filter with scale $\sigma = 1$ pixel.

3.4. Initial Candidate Detection

The initial candidate detection steps described in this section are illustrated in figure 3. In the descriptions below a number of empirically decided threshold values are mentioned. These threshold values were determined during an initial development period on a dataset completely independent of the training and test sets used in this work. Actual values used for thresholds are listed in figure 4. An example of nodule candidates found in a test scan can be seen in figure 5.

3.4.1. Seed Point Detection

Once the SI and CV values for the image are known, a set of seed points is established by the thresholding of these values according to empirically decided limits SI_{upper}, SI_{lower}, CV_{upper} and CV_{lower}. Voxels which have both SI and CV within the thresholds are selected as seeds. For locations within 5 voxels of the pleural wall a slightly lower value of SI_{lower} is used in order to increase the size of clusters found in the pleural region. Such clusters are typically smaller than non-pleural examples since pleural nodules do not present as much surface area for examination, and in addition the SI and CV values in their region may be affected by the topology of nearby pulmonary walls.

---

*The definition of CV used here excludes the scaling constant of $\sqrt{2}$ used by Koenderink[21] which serves only to enforce unit curvedness on the unit sphere.
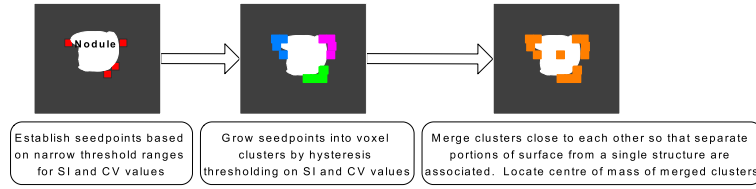
Figure 3. Visualisation of Detection of Initial Candidates using SI and CV values

3.4.2. Cluster Formation

The seedpoints are now expanded to form clusters of voxels of interest. The expansion is based on hysteresis thresholding[19] using additional thresholds $SI_{\text{above\_upper}}$, $SI_{\text{below\_lower}}$, $CV_{\text{above\_upper}}$ and $CV_{\text{below\_lower}}$. These are empirically chosen with

$$SI_{\text{above\_upper}} \geq SI_{\text{upper}}, \quad SI_{\text{below\_lower}} \leq SI_{\text{lower}}$$
$$CV_{\text{above\_upper}} \geq CV_{\text{upper}}, \quad CV_{\text{below\_lower}} \leq CV_{\text{lower}}$$

The final cluster contains only voxels whose SI and CV values fall within the broader threshold range and which can be connected to a seedpoint by a chain of other such voxels.

It should be noted that for a perfectly spherical structure, the voxels in the final cluster lie in the region of the blurred surface of the sphere. The centre of mass of the cluster is taken to be the point of interest in nodule detection. References to cluster 'location' throughout this work imply the centre of mass of the related set of voxels.

A cluster whose original seedpoint lies within 5 voxels of the pulmonary wall is considered to be a pleural candidate. A lower minimum volume t_{vol} is defined for pleural clusters as explained in section 3.4.1. At this stage clusters whose volume is below a pre-determined threshold t_{vol} are discarded as their inclusion in the remaining processing steps was found to be extremely costly and more likely to introduce false-positives.

Seed Detection		Hysteresis Thresholding		Cluster Size	
SI_{lower} (pleural)	0.9	$SI_{\text{below\_lower}}$	0.7	t_{vol} (pleural)	15
SI_{lower} (non-pleural)	0.8	$SI_{\text{above\_upper}}$	1.0	t_{vol} (non-pleural)	4
SI_{upper}	1.0	$CV_{\text{below\_lower}}$	0.2		
CV_{lower}	0.3	$CV_{\text{above\_upper}}$	1.3		
CV_{upper}	1.0				

Figure 4. Thresholds used during nodule detection process
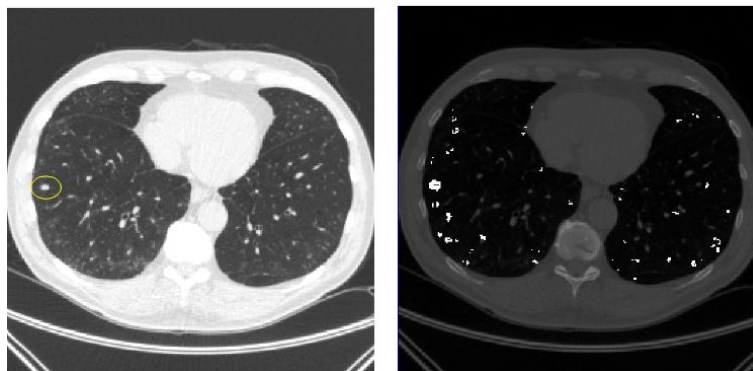

Figure 5. An example of clusters found in initial candidate detection as seen in a singe slice. The true nodule is circled in the leftmost image. Cluster voxels are shown in bright white in the image on the right.

3.4.3. Cluster Merging

At this stage a large number of clusters have been detected. Each one represents a region of surface in the image and it is reasonable to suppose that a true structure such as a nodule may have more than one cluster representing it. Except in the case where the nodule is extremely large and unusually shaped these clusters will lie in close proximity to each other. Clusters with locations within 3 voxels of each other are therefore recursively merged until no more merges can be performed, and the procedure is then repeated for clusters within 7 voxels of each other.

3.5. False Positive Reduction by kNN-Classification

In order to reduce the numbers of false-positives kNN-classification[18] is now used to assign each candidate structure to a final class of 'nodule' or 'non-nodule'. Two individual classification steps take place, as described in sections 3.5.4 and 3.5.5.

The initial classification step uses a small number of relatively simple features to quickly reduce the most obviously incorrect candidates. The second classifier employs more features of higher complexity in order to classify the more ambiguous remaining candidates as accurately as possible.

3.5.1. Training Data

The scans used in building the training sets form part of the Nelson study data and include patients scanned in Utrecht University hospital in June, July and August of 2005. In total 698 scans with 1462 annotated nodules are used in the creation of two training sets for the kNN classifiers to be used in false-positive reduction. The training datasets are built using the system of initial nodule candidate detection described in section 3.4 and classifying a detected candidate as 'nodule' if its location is within 7 voxels of an annotation marked by the radiologist and 'non-nodule' otherwise. Training data for the first-step classifier was constructed initially and used to train a temporary kNN-classifier. This trained first-step classifier was then used to produce training data for a kNN-classifier for use in the second-step. In this way the candidates used for training emulate as closely as possible the characteristics of the candidates that will be encountered during the classification phase. Figure 6 illustrates the procedure used to generate training sets in full.

It is clear that datasets generated in this manner will not be balanced in terms of the number of items in each class. Since the performance of the kNN-classifier is optimal where the training sets have reasonably well-balanced class sizes[18] it was decided to reduce the size of the 'non-nodule' class so that it is closer in size to the 'nodule' class. In order to ensure that the distribution of the samples in the dataset is not unduly altered a data condensation method[22] is employed to reduce the number of false-positive samples.

3.5.2. Test Data

The samples to be classified by the trained kNN-Classifier are feature vectors of nodule candidates detected in Nelson study subjects scanned in Utrecht University in October 2005. Patients who had earlier scans included in the training phase were excluded from the test set to ensure complete independence of the training and test data. A random selection of 142 scans containing a total of 268 nodules was chosen for testing the algorithm.

3.5.3. Feature Extraction

This section describes all 83 features which were initially considered as being potentially useful in distinguishing nodules from false-positive candidates. As described in sections 3.5.4 and 3.5.5 a 'Sequential Forward Floating Selection' (SFFS)[23] feature selection procedure was employed in both classification steps to choose the most useful features. The feature-set was chosen to maximise the area under the ROC curve achieved by leave-one-out training and testing on the training data. The descriptions below make reference to the 'radius' of the cluster, *clusterRadius*. This value is determined by summing the maximum diameter of the cluster in each of the directions X, Y and Z, dividing by 3 to get an average diameter, and by 2 to get an average radius.

Features of the voxel cluster A number of basic features are calculated relating to the size and shape of the cluster of voxels which constitutes the nodule candidate.

- The volume (number of voxels), *vol* of the cluster.
- The maximum dimension *max_dim* from the X, Y or Z directions.
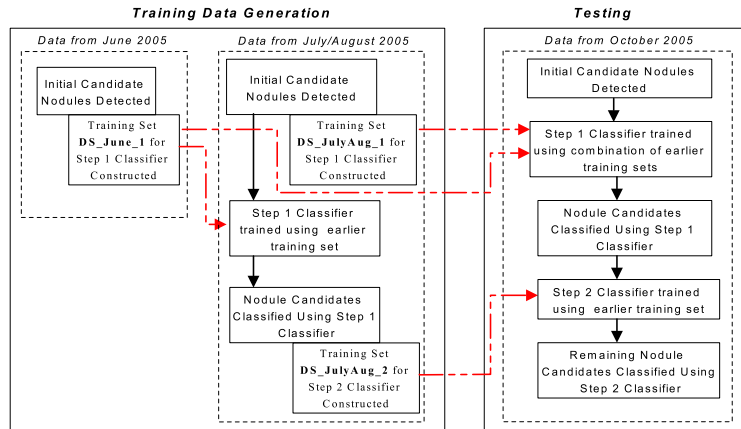- The minumum dimension *min_dim* from the X, Y or Z directions.

Figure 6. The procedure of generating and using training data for kNN-classification

- $Compactness_1$, defined by $\frac{vol}{(dim_x)(dim_y)(dim_z)}$, where dim_i is the size of the component in dimension i.
- $Compactness_2$, defined by $\frac{vol}{max\_dim^3}$
- The ratio max_dim/min_dim
- The ratio max_dim/med_dim, where med_dim is the second largest dimension from the X, Y or Z directions.
- Sphericity of the cluster which is defined as follows: $\frac{numVoxelsFromCluster}{totalNumVoxels}$ where $totalNumVoxels$ is the number of voxels in a spherical kernel with radius $clusterRadius$ and $numVoxelsFromCluster$ is the number of voxels from the cluster which fall within such a spherical kernel centred at the candidate location. See figure 8(a).

Features of grey-values in the neighbourhood of the candidate Spherical kernels of various sizes are defined, centred at the candidate location. Radii of 1, 3, and $clusterRadius$ were used. Grey-values within each defined kernel are examined and the following features are calculated over the voxels encountered:

- Average grey-value.
- Minimum grey-value.
- Maximum grey-value.
- Median grey-value.
- Standard Deviation of grey-values.

The use of these features is designed to eliminate false-positives which are located in a region that is insufficiently bright to represent a true nodule.

Features of SI and CV in the neighbourhood of the candidate Spherical kernels of various sizes are defined, centred at the candidate location. Radii of 3 and $clusterRadius$ were used for the kernel sizes. The kernel SI and CV values within each defined kernel are examined and the following features are calculated over the voxels encountered:

- Average value of both SI and CV.
- Minimum value of both SI and CV.
- Maximum value of both SI and CV.
- Median value of both SI and CV.
- Standard Deviation of values for both SI and CV.

Features of Intensity Gradient Orientation and Magnitude in the neighbourhood of the candidate
The distribution of grey-levels across cross-sections of a true nodule is usually approximately similar to a

central-symmetric Gaussian distribution. This implies the relevance of gradient features in distinguishing true nodules from false-positives. Consider, for example, a spherical surface, where the sphere is centred at a nodule location and its surface lies within the nodule structure(See figure 7(a)). The gradient magnitude at all points on the sphere surface should be approximately the same, and the standard deviation of these magnitudes is therefore close to zero. Furthermore, the orientation of the gradient at these points should be approximately the same as the radial direction.

The calculation of gradient features proceeds as follows. Two spherical surfaces centred at the nodule candidate location are defined, with radii of 3 and $clusterRadius$ respectively. A number of points is randomly chosen on the surface of each sphere (30 points on the first sphere and fifty points on the second, since it is likely to be larger) and gradient magnitude and orientation are calculated at each point. Gradient orientation is defined by the normalised component of the gradient vector which lies in the radial direction, i.e. $V_{G_{Radial}} = \frac{V_R \cdot V_G}{|V_R||V_G|}$ where V_R is the radial vector and V_G is the gradient vector (see figure 7(b)).

For each of the two defined spheres the following statistics are now calculated based on the set of random points selected on the sphere surface:

- The average (μ), minimum (min), maximum (max), median (med) and standard deviation (σ) of the set of values for both gradient orientation and gradient magnitude.
- The ratio max/min for both gradient orientation and gradient magnitude.
- The coefficient of variation, ($\frac{\sigma}{\mu}$), for both gradient orientation and gradient magnitude.
- The ratio σ/med for both gradient orientation and gradient magnitude.
- The ratio med/μ for both gradient orientation and gradient magnitude.
- The ratio med/max for both gradient orientation and gradient magnitude.

(a) Gradient Properties under perfect conditions   (b) Radial Gradient
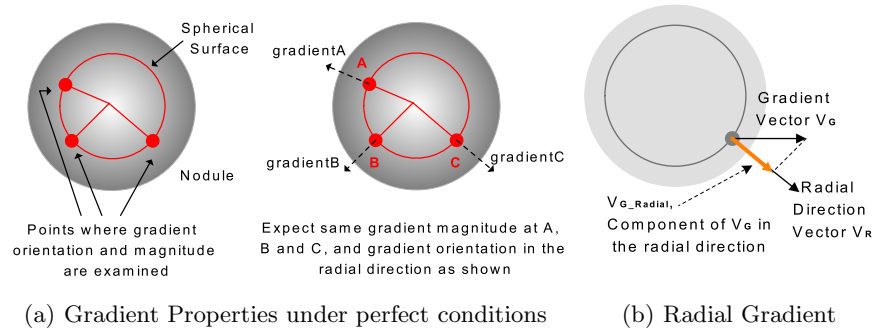
Figure 7. Nodule Gradient Properties

3.5.4. First Classification

After initial nodule candidate detection an average of 703 structures per scan are established for classification. The computation of all features listed in the previous section for each of these objects is not an efficient way to proceed since many of them are easily identifiable as false-positives based on a much smaller number of simpler features. For this preliminary classification step we therefore examine only features of the voxel cluster and features of grey-values as described in section 3.5.3.

A kNN-classifier[18] is employed and supplied with the training set of 7820 samples. The chosen value of k is the (odd) square-root of the number of samples, in this case 89. A 'Sequential Forward Floating Selection' (SFFS) feature-selection procedure[23] is applied to the training set to determine the most discriminatory features among them before training the kNN-classifier. The feature-selection is carried out using leave-one-out training and testing on the training dataset. In this way 8 features were chosen for the first step classifier from the voxel cluster and grey-value features which were calculated. Figure 8(b) shows the selected features. The training dataset is now reduced to contain only the relevant features and the first-step classifier is trained.

(a) Sphericity Definition
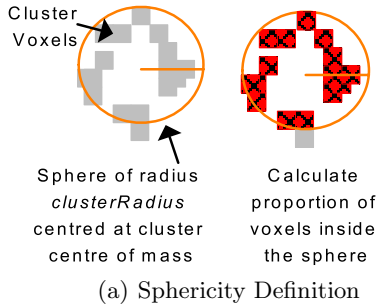
Grey Value Features over a spherical volume	Sphere Radius
Median	1
Average Std Dev σ	3
Median Std Dev σ	*cluster Radius*

Cluster Features
Cluster Volume Sphericity Ratio Max_Dim/Med_Dim

(b) Features selected for use by first step classifier

Figure 8. The definition of sphericity and the features used in the first step classifier

Next the candidates awaiting classification are examined and the same 8 features are calculated for each of them to form the test-data. The trained kNN-classifier is now used to classify each candidate as 'nodule' or 'non-nodule' based on its feature data.

After first step classification as described in this section an average of only 84 of the initial 703 candidates per scan remain for further classification. The kNN-Classifier achieves 94% sensitivity and 88% specificity in classifying the candidates presented to it. The FROC curve for the first classification procedure is shown in figure 10(a).

3.5.5. Final Classification

The final classification step uses a second kNN-classifier to further reduce the number of false-positives among the remaining candidates. At this stage there are an average of 84 candidates per scan awaiting final classification. The training set used for the kNN-classifier consists of 3987 samples for which all features described in section 3.5.3 have been calculated. Each sample is therefore 83-dimensional and an SFFS feature selector is used as described in section 3.5.4 to reduce the sample dimensionality and retain the set of features which give optimal results on leave-one-out testing within the training set only. The value of k for the second kNN classifier is set once again as the square-root of the number of samples, which is 63 in this case. The 22 features retained by the feature selection system are shown in figure 9:

Gradient Orientation Features over points from a sphere surface	Number of points	Sphere Radius
Coefficient of Variation Ratio med/μ	30	3
Coefficient of Variation Ratio med/μ	50	*cluster Radius*

CV Features over spherical volume	Sphere Radius
Median Std Dev σ	3
Median Average	*clusterRadius*

Gradient Magnitude Features over points from a sphere surface	Number of points	Sphere Radius
Maximum	30	3

SI Features over spherical volume	Sphere Radius
Average Std Dev σ Median	3

Cluster Features
Cluster Volume Sphericity

Grey Value Features over spherical volume	Sphere Radius
Maximum	1
Median Std Dev σ Minimum Maximum	3
Std Dev σ Minimum Maximum	*cluster Radius*

Figure 9. Features selected for use by the final Classifier

The training set data samples are now reduced to contain only the selected features and the kNN classifier is trained. Next the candidates awaiting classification are analysed and their features calculated, creating a sample point in 22-Dimensional dataspace for each one. The trained classifier is then used to classify each candidate as 'nodule' or 'non-nodule'. The final classification step reduces the total number of detections from 84 to an average of just 10 per scan. A sensitivity of 90% is achieved by the classifier with a specificity level also of 90%. The FROC curve for the final classification is shown in figure 10(b).

4. RESULTS

The algorithm was tested on 142 randomly chosen scans which contained a total of 268 nodules, 66 of which were pleural (see figure 11(b)). The average diameter of the nodules ranged between 2mm and 14mm. The range of
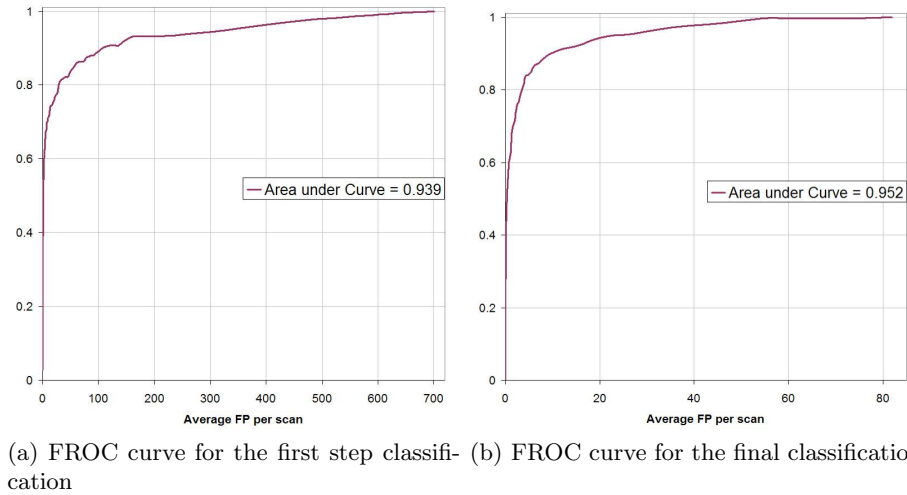
(a) FROC curve for the first step classifi-   (b) FROC curve for the final classification
cation

Figure 10. FROC Analysis for the two classification steps using the full dataset of 142 scans.

nodule sizes is shown in figure 11(a). Without placing any further restrictions on the scans to be included in the testing phase the algorithm achieved a sensitivity of 84% (224 of the 268 nodules) with an average of 8.2 false-positive detections per scan. Table 1 summarises the results of the scheme at each stage of the detection procedure.

Initial Candidate Detection		After first classification		After second classification	
% Sensitivity	False-Positives	% Sensitivity	False-Positives	% Sensitivity	False-Positives
99%	703	93%	82	84%	8.2

Table 1. Average results per scan for each stage of the Nodule Detection Procedure.

Of the nodules which were missed by the algorithm, 28 were observed to be pleural indicating a significantly better overall performance on non-pleural nodule detection (see figure 11(b)). Considering pleural and non-pleural nodules separately a sensitivity of 92% is found in the case of non-pleural structures while for pleural nodule detection the sensitivity is 58%.

The false-positive detections were commonly located at vessel junctions or where vessels curved sharply, in the complex topology of the mediastinum or at minor undulations on the pleural wall.

Figure 11(a) illustrates that 84% of the nodules which the system failed to detect were below 5mm in diameter. Nodules below this size are extremely unlikely to be diagnosed as malignant.[24] In addition all but 2 of 17 nodules whose shape was characterised as lobulated or spiculated were successfully detected. These nodule types are more likely to be diagnosed as malignant than their smooth-edged counterparts,[25] making their detection extremely important although often problematic for CAD systems. (The remainder of the nodules had shapes characterised as 'compact')

It is believed that the system results are affected by ommissions and inaccuracies in the annotations used as ground-truth, since these are the work of one radiologist in a single reading. A second expert examination of these annotations is likely to result in an altered ground-truth and hence an improvement to the system results.[11] In addition, a number of the scans included in the test data used in this work display signs of interstitial disease. It was decided in the absence of expert diagnosis to retain them in the test set although many authors deliberately remove such scans from their studies. Scans displaying interstitial disease are likely to provide a large number of false-positives due to the high intensity patterns associated with the disease which are seen in the lung region. In addition these textural abnormalities could occlude true nodules making them more difficult to detect. It may be concluded that the exclusion of scans exhibiting interstitial disease would be likely to improve the average sensitivity and specificity achieved by our algorithm.

Although a direct comparison with other systems is not possible due to the disparity in the test data used,

it is worthwhile to consider the results achieved in the context of previous system performances. In particular a commercial system approved for use by the FDA[6] is used for comparison in order to gauge the potential usefulness of the system developed in this work. This commercial system achieved a sensitivity of 60% at 1.56 FP per scan[6] on a data set which was chosen to exclude scans exhibiting lung disease, masses larger than 2cm, multiple nodules (more than 6), large pleural effusions, non-solid or partly-solid nodules. The test data consisted of 78 scans with nodule diameters between 4mm and 15.4mm. In order to make a comparison with the algorithm developed in this work, the sensitivity rate for our system at 1.56 FP per scan was extracted from the FROC data plotted in figure 10(b). The final kNN classifier is found to be capable of correctly identifying 69% of the true nodules presented to it with an average of 1.56 false detections per scan. Taken in conjunction with the sensitivity of the initial detection scheme (99%) and of the first kNN classification (94%), an overall system performance of 64.2% sensitivity at 1.56 FP per scan is obtained. The system developed here therefore outperforms that which was tested in [6] without placing any restrictions on the randomly chosen test data.
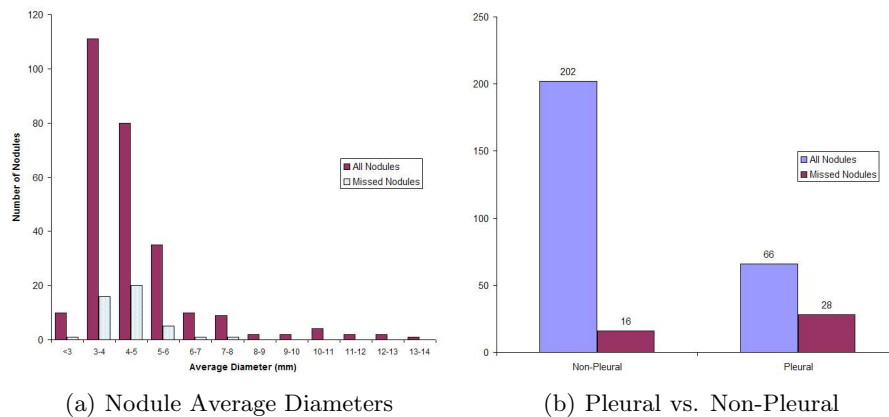

(a) Nodule Average Diameters   (b) Pleural vs. Non-Pleural

Figure 11. Statistics on the average diameters and the location (pleural or non-pleural) of the nodules from the scans used in testing.

5. CONCLUSION

A system has been developed for automated lung nodule detection in CT based on local image features. The system has been trained and tested on a large and diverse database of scans and performs well on a broad range of nodule types and sizes, particularly those which may be more likely to be malignant. The overall sensitivity and specificity achieved compare favourably in general with those reported from other schemes.[6–14] This is particularly true when it is considered that the database on which our system was tested is larger than in any of the other cases and not manually restricted in any way. Specifically, the randomly chosen scans included patients with interstitial lung disease, with up to 14 annotated nodules in a single scan and without limitation on the nodule sizes present in the data. Comparison with an FDA-approved commercial system gives favourable results with the developed algorithm showing improved sensitivity at the same false-positive rate as that achieved by the authorised system. The data set used in testing our algorithm was almost twice as large as that used in the tests for the commercial system and included a broader range of nodule sizes and types without the exclusion of non-solid or partly solid nodules or of scans exhibiting lung disease. The system developed therefore demonstrates a unique robustness in peformance over a broad spectrum of cases representative of those which would be encountered in a screening programme.

REFERENCES

1. A. C. Society, "Cancer facts and figures, 2006.".
2. C. Henschke, D. I. McCauley, D. F. Yankelevitz, D. P. Naidich, G. McGuinness, O. S. Miettinen, D. M. Libby, M. Pasmantier, J. Koizumi, N. K. Altorki, and J. P. Smith, "Early Lung Cancer Action Project: Overall design and findings from baseline screening," *The Lancet* **354**, pp. 99–105, 1999.

3. R. Kakinuma, H. Ohmatsu, M. Kaneko, K. Eguchi, T. Naruke, K. Nagai, Y. Nishiwaki, A. Suzuki, and N. Moriyama, "Detection failures in spiral CT screening for lung cancer: Analysis of CT findings," *Radiology* **212**, pp. 61–66, 1999.

4. J. Gurney, "Missed lung cancer at CT: Imaging findings in nine patients," *Radiology* **199**, pp. 117–122, 1996.

5. G. D. Rubin, J. K. Lyo, D. S. Paik, A. J. Sherbondy, L. C. Chow, A. N. Leung, R. Mindelzun, P. K. Schraedley-Desmond, S. E. Zinck, D. P. Naidich, and S. Napel, "Pulmonary nodules on multi-detector row CT scans: Performance comparison of radiologists and computer-aided detection," *Radiology* **234**, pp. 274–283, 2005.

6. I. Lee, G. Gamsu, J. Czum, R. J. N. Wu, and S. Chakrapani, "Lung nodule detection on chest CT: Evaluation of a computer-aided detection (CAD) system," *Korean Journal of Radiology* **6(2)**, pp. 89–93, 2005.

7. S. Armato, M. Giger, and H. MacMahon, "Automated detection of lung nodules in CT scans: Preliminary results," *Medical Physics* **8**, pp. 1552–1561, 2001.

8. H. Arimura, S. Katsuragawa, K. Suzuki, and et al., "Computerized scheme for automated detection of lung nodules in low-dose computed tomography images for lung cancer screening," *Academic Radiology* **11(6)**, pp. 617–629, 2004.

9. Q. Li, S. Sone, and K. Doi, "Selective enhancement filters for nodules, vessels and airway walls in two- and three-dimensional CT scans," *Medical Physics* **30(8)**, pp. 2040–2051, 2003.

10. D. Paik, C.F.Beaulieu, G. Rubin, B. Acar, R. B. J. Jr, J. Yee, J. Dey, and S. Napel, "Surface Normal Overlap: A computer-aided detection algorithm with application to colonic polyps and lung nodules in helical CT," *IEEE Transactions on Medical Imaging* **23(6)**, pp. 661–675, 2004.

11. X. Zhang, G. McLennan, E. Hoffman, and M. Sonka, "Automated detection of small-size pulmonary nodules based on helical CT images," in *Proceedings of Information Processing in Medical Imaging:19th International Conference*, *Lecture Notes in Computer Science*, p. 664, Springer-Verlag, 2005.

12. M. Gurcan, B. Sahiner, N. Petrick, H. Chan, E. Kazerooni, P. Cascade, and L. Hadjiiski, "Lung nodule detection on thoracic computed tomography images: Preliminary evaluation of a computer-aided diagnosis system," *Medical Physics* **29**, pp. 2552–2558, 2002.

13. Z. Ge, B. Sahiner, H. Chan, L. Hadjiiski, and et al., "Computer-aided detection of lung nodules: False positive reduction using a 3D gradient field method," in *Medical Imaging 2004: Image Processing*, M. Sonka and J. M. Fitzpatrick, eds., *Proceedings of SPIE* **5370**, pp. 1076–1082, 2004.

14. B. Sahiner, Z. Ge, H. Chan, L. Hadjiiski, and et al., "False-positive reduction using Hessian features in computer-aided detection of pulmonary nodules on thoracic CT images," in *Medical Imaging 2005: Image Processing*, J. Fitzpatrick and J. Reinhardt, eds., *Proceedings of SPIE* **5747**, pp. 790–795, 2005.

15. Y. Lee, T. Hara, H. Fujita, S. Itoh, and T. Ishigaki, "Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique," *IEEE Transactions on Medical Imaging* **20(7)**, pp. 595–604, 2001.

16. A. Farag, A. El-Baz, G. Gimel'farb, R. Falk, and S. Hushek, "Automatic detection and recognition of lung abnormalities in helical CT images using deformable templates," in *Medical Image Computing and Computer-Assisted Intervention*, C. Barillot, D. Haynor, and P. Hellier, eds., *Lecture Notes in Computer Science* **3217**, pp. 856–864, Springer-Verlag, 2004.

17. "http://www.r2tech.com/main/company/news.php.".

18. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley Interscience, 2000.

19. J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **8**, pp. 679–698, 1986.

20. S. Hu, E. Hoffman, and J. Reinhardt, "Automatic lung segmentation for accurate quantitation of volumetric X-Ray CT images," *IEEE Transactions on Medical Imaging* **20(6)**, pp. 490–498, 2001.

21. J. J. Koenderink, *Solid Shape*, MIT Press, Cambridge, MA, 1990.

22. P. Mitra, C. Murthy, and S. Pal, "Density-based multiscale data condensation," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24(6)**, pp. 734–747, 2002.

23. P. Pudil, J. Novovicova, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Letters* **15(11)**, pp. 1119–1125, 1994.

24. C. I. Henschke, D. F. Yankelevitz, D. P. Naidich, D. I. McCauley, G. McGuinness, D. M. Libby, J. P. Smith, M. W. Pasmantier, and O. S. Miettinen, "CT screening for lung cancer: Suspiciousness of nodules according to size on baseline scans," *Radiology* **231**, pp. 164–168, 2004.

25. J. Gurney, "Determining the likelihood of malignancy in solitary pulmonary nodules with Bayesian analysis. Part I. Theory," *Radiology* **186(2)**, pp. 405–413, 1993.