# IMPROVED CLASSIFICATION OF PULMONARY NODULES BY AUTOMATED DETECTION OF BENIGN SUBPLEURAL LYMPH NODES

*M.A.J. Klik[1], E.M. v. Rikxoort[1], J. F. Peters[2], H.A. Gietema[3], M. Prokop[3], and B. v. Ginneken[1]*

[1]Image Sciences Institute, Utrecht University, Heidelberglaan 100, 3584 CX Utrecht, Netherlands
[2]Healthcare IT - Advanced Development, Philips Medical Systems, Veenpluis 4-6, 5680 DA Best
[3]University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, Netherlands

## ABSTRACT

A new computer algorithm is presented to distinguish a special, most probably benign, subclass of lung nodules called perifissural opacities (PFO's), from potentially malignant nodules. The method focuses on the quantification of two characteristic properties of PFO's, namely the typical flattened surface of the nodule and its attachment to plate-like structures in the direct neighborhood of the nodule (the lung fissures). For the detection of fissures in the proximity of the nodule, an analysis based on the eigenvalues of the Hessian matrix has been developed. Further processing with a voxel grouping algorithm is shown to substantially improve the results of the fissure detection. Through a comparison of Hough transforms of the nodule boundary and the detected fissure voxels, features are constructed that enable a reliable separation of benign PFO from other lesions.

## 1. INTRODUCTION

The use of modern multi-row CT (MDRCT) technology allows for fast, high resolution (sub-millimeter) near isotropic acquisition of the lungs. MDRCT enables the detection of small pulmonary nodules which can lead to early detection of lung cancer. The 5 year survival rate of patients with completely resected stage 1A or 1B non-small lung cancer is 65 to 85 percent [1], as compared to only 10 to 15 percent for other patients diagnosed with lung cancer. Several lung cancer screening studies are underway to investigate whether mortality from lung cancer can be reduce by screening. The Dutch Nelson screening program is such a study where low-dose state-of-the art CT scanning technology is utilized to investigate if mortality is reduced compared to a control group. In four years, 10000 participants will be scanned three times with 16-slice low-dose CT. The data presented in the current paper is taken from this screening study.

It turns out that a significant proportion of CT scans from randomly selected high-risk individuals (typical for screening) contain medium sized nodules with volumes in the range from 50 to 500 mm$^3$. Because follow-up procedures to determine malignancy of the nodule are invasive and induce risk for the patient, an accurate classification of such nodules into benign lesions versus malignant tumors is a crucially important task in screening studies as well as in chest radiology in general.

Recent clinical research [2] has shown that up to a third of pulmonary nodules, so called perifissural opacities (PFO's), most probably represent sub pleural lymph nodes. PFO's are characterized by their connection to one or more fissures (major/accessory/pleura) in the lungs (fissures are thin sheets of tissue that separate the compartments of the lungs) and the typical flattened shape that is the result of these connections. An example of a PFO is shown in Fig.1A. The two major characteristics for a PFO can clearly be observed in this image; the connection to a fissure is obvious and the nodule has a typical flattened shape on the side that is connected to the fissure. In Fig.1B, a non PFO nodule is shown for comparison. No connection with a fissure or flattened sides can be observed for this nodule. We present a novel classification system to distinguish these PFO's from potentially malignant nodules. Such a classification system is very important for reduction of the false-positive rate and has the potential to increase diagnostic accuracy. This paper presents a CAD system that is based on a plate-detection algorithm applied to the nodule surface (to detect flat sides) and to the surroundings of the nodule (to detect fissures). We will show that by calculating the Hough transform of the image of the segmented nodule surface and comparing that with the Hough transform of the detected plate-like voxels in the neighborhood of the nodule, a reliable separation of PFO's from other, potentially malignant nodules can be made.

## 2. METHOD

### 2.1. Data set

The investigated pulmonary nodules are taken from low-dose spiral CT images (16 x 0.75 mm collimation, 30 mAs) that were acquired with a Philips scanner type Mx8000IDT, at full inspiration. 3D volumes of interest (60 x 60 x 60 mm) around 284 nodules with a volume between 50 and 500 mm$^3$, detected in the exams of 221 subjects, were presented to a chest radiologist, along with the complete CT image. The

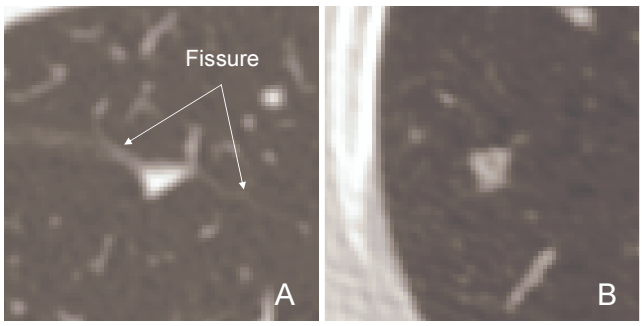


**Fig. 1**. Two types of nodules. A) a PFO that is attached to a fissure. B) non PFO nodule.

radiologist classified each nodule as PFO or not a PFO. In addition, the nodules were categorized according to their shape and their spatial relationship to the fissures in the lungs.

### 2.2. Nodule segmentation

With our computerized scheme, a nodule was automatically segmented using a segmentation algorithm developed by Philips Research Laboratories Hamburg [3]. The seed-point chosen for the algorithm was the point indicating the location of the nodule as clicked by the radiologist. The total volume of interest (VOI) consists of a cubic area 60 x 60 x 60 mm in size around the center of mass of the nodule. From the binary segmentation image the location of voxels at the surface of the nodule are determined. These voxels are used in the algorithm described in paragraph 2.4, that enables the detection of flattened sides on the nodule.

### 2.3. Fissure detection

Detection of fissures inside the VOI is based on the ratio of the first ($h_1$) and second ($h_2$) eigenvalues of the Hessian matrix, ordered in decreasing magnitude as described in [4]. A "plateness" value P is computed for each voxel from the eigenvalues:

$$\begin{aligned} P &= \frac{|h_1| - |h_2|}{|h_1| + |h_2|} \cdot e^{\frac{-(I-\mu)^2}{2\sigma^2}}, \text{ for } h_1 < 0 \\ P &= 0, \text{ otherwise,} \end{aligned} \quad (1)$$

with $I$ the intensity of the voxel, $\mu$ the average intensity for a fissure in the lungs and $\sigma$ the standard deviation of fissure voxel intensity. The term $\exp[(I-\mu)^2/2\sigma^2]$ takes the expected intensity value of a fissure voxel into account. The second order derivatives of intensity, used to construct the Hessian matrix, are calculated using the Gaussian scale-space technique. For plate-like structures, a local maximum is expected in the intensity (first eigenvalue of Hessian $< 0$), while in directions perpendicular to that, the eigenvalues are expected to be relatively small.

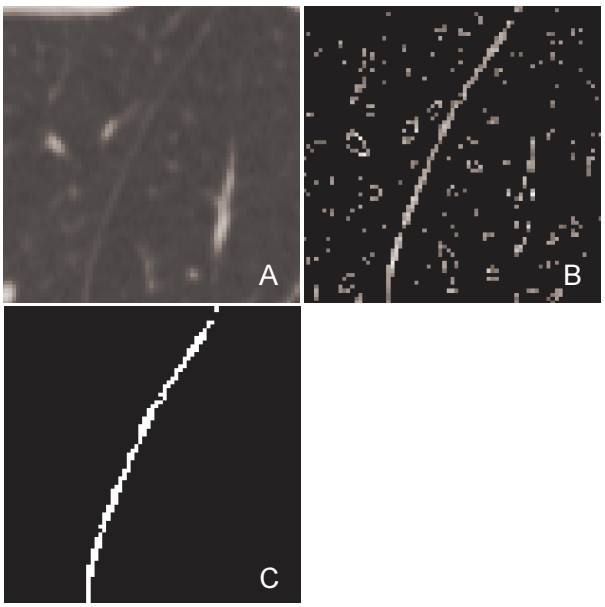


**Fig. 2**. Algorithm to detect a fissure around the nodule. A) the original volume of interest (no nodule is visible). B) an image that depicts the value of parameter $P$ ("plateness") for each voxel in the original slide shown. C) Binary image after processing with the grouping algorithm.

In Fig.2A, one slice of a VOI is shown alongside the values of $P$ as calculated for this VOI (Fig.2B). A threshold of $P \geq 0.2$ is used, so smaller values are not displayed in the image. It can be seen that besides the fissure that we aim to detect, additional structures show up in this "plateness" image that are only plate-like on a scale typically of the order of a few millimeters. Increasing the threshold to larger $P$, or increasing the scale of the Gaussian derivatives used to compute the Hessian matrix, removes many of those responses, but decreases the sensitivity to the real fissures. In order to enable fissure detection without sacrificing sensitivity, we developed an algorithm that groups neighboring detected voxels based on the direction of the first eigenvector of the Hessian and their likelihood of constituting a plane, inspired by the work described in [5]. The algorithm is explained in Fig.3. Vectors $\overrightarrow{\mathbf{v^1}}$ and $\overrightarrow{\mathbf{v^2}}$ denote the normalized first Hessian eigenvectors of points $\mathbf{p^1}$ and $\mathbf{p^2}$, and $\overrightarrow{\mathbf{w^1}}$ is the difference vector between source point $\mathbf{p^1}$, and candidate point $\mathbf{p^2}$. We demand that $\overrightarrow{\mathbf{v^1}} \cdot \overrightarrow{\mathbf{v^2}}$ is larger than a threshold $C_{pa}$, and that $\overrightarrow{\mathbf{v^1}} \times \overrightarrow{\mathbf{w^1}}$ is larger than a threshold $C_{ip}$. In that case, candidate point $\mathbf{p^2}$ is added to the end of the group that contains $\mathbf{p^1}$. Point $\mathbf{p^2}$ is now removed as a candidate point. If all candidate points in the vicinity of $\mathbf{p^1}$ are evaluated, the next point in the group is set as the source point. This means that also

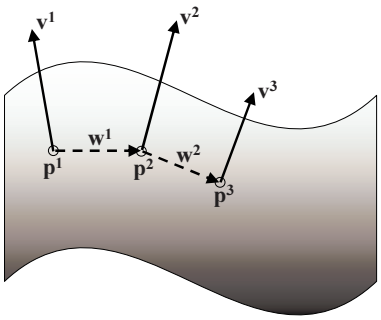


**Fig. 3**. The grouping algorithm connects voxels on basis of the direction of the first eigenvector of the Hessian matrix, and the likelihood of the grouped voxels of constituting a plane.

point $\mathbf{p^2}$ will eventually be considered as source point. All points within a certain distance $d$ from the "source" point are considered as candidates. By restricting the minimum size of groups of points, small, locally "plate-like" structures can be discarded. Fig.2C depicts a binary image of all the grouped points which resulted from our grouping algorithm with parameters $C_{ip} = C_{pa} = 0.985$ and a maximum distance $d$ of 3.1. The minimum group size was chosen to be 200 voxels. It can be seen that only those voxels remain which correspond to the fissure in the VOI.

### 2.4. Hough transform representation

As indicated before, there are two major characteristics for PFO's; their connection to one or more fissures (major, accessory, or pleura) in the lungs and the typical flattened shape that is the result of these connections. This connection implies that the direction of a plane fitted to the flattened side of the nodule should be similar to the direction of the detected plate-like voxels in the nodule neighborhood. We quantify this relationship by calculating the Hough transform [6] of the binary image of the detected plate-like voxels (Fig.2C) as well as the Hough transform of the surface voxels of the binary segmented nodule image, shown in Fig.4. Using the seed-point as our origin, we calculate for each voxel in the binary image all possible planes that run through this voxel. Since a plane in 3D can be described by its shortest distance from the origin $l$, and two angles $\alpha$ and $\beta$, indicating the angle with the z-axis and x-axis respectively, the Hough transform of such a plane is a point in 3 dimensional Hough-space. For each voxel set in the binary image, all possible combinations of $l$, $\alpha$ and $\beta$ are calculated and these solutions are accumulated in the Hough transform matrix. In Fig.5, the Hough transform accumulators that are calculated for the nodule boundary image (A) and the plate voxels image (B) are depicted. Slices are shown for $l = 2$. Although the average intensity of the two plots is very different because of different numbers of processed voxels, it can clearly be seen that maxima occur at

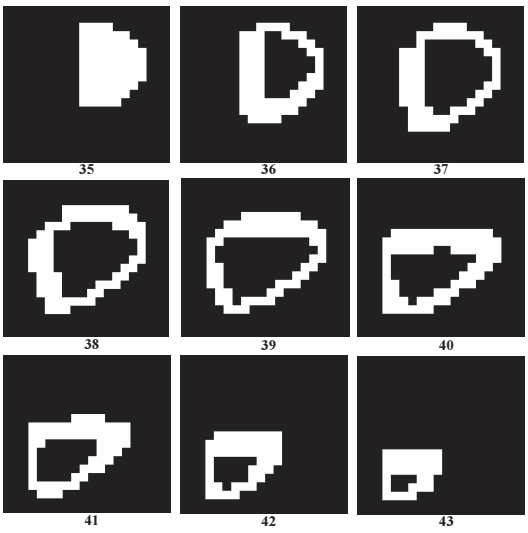


**Fig. 4**. Various slices of a (3D) segmented lung nodule boundary. It can be seen that the nodule is flattened on one side.

almost identical locations. This reflects the fact that the flat part of the nodule boundary coincides with a plane through many of the plate-like voxels, indicating a connection of the nodule with a fissure.

### 2.5. Features

The intensity of the Hough transform of nodule surface voxels depends on the actual number of voxels used in the calculation. To correct for this, the transform is divided by this amount. This normalized image is then multiplied with the Hough transform calculated for the plate-like voxels, as shown in Fig.5C. The resulting multiplication image will have maxima at the coordinates ($l$, $\alpha$, $\beta$) where the intensity of the two Hough transforms both have high values. We take the intensities of the three largest maxima found in the multiplication image as features $f_1$, $f_2$, and $f_3$, the three largest maxima in the Hough transform of the nodule surface voxels as features $f_4$ to $f_6$, and the three maxima in the Hough transform of the fissure voxels as features $f_7$ to $f_9$. Three additional features $f_{10}$ to $f_{12}$ are added which are a function of the other features: $f_{10} = f_3/f_1$, $f_{11} = f_6/f_4$, and $f_{12} = f_9/f_7$. These features are added to discriminate between Hough transforms images with a large number of maxima of approximately similar intensity from transforms with just one or two maxima.

## 3. RESULTS

A total of 284 nodules were identified as the reference standard. Of these, 35 percent (99) were categorized as a PFO by the radiologist. Several classifiers were used to train the system with the normalized features mentioned above; a $k$NN

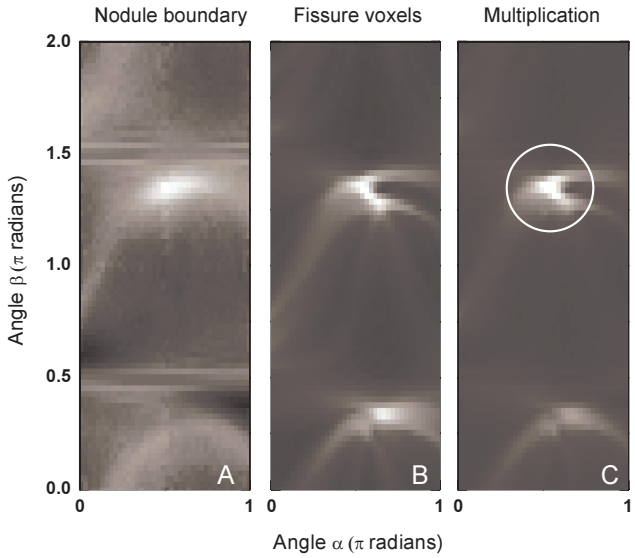


**Fig. 5**. Hough transforms of A) the segmented nodule boundary and B) the detected fissure voxels. C) Multiplication image of A and B.

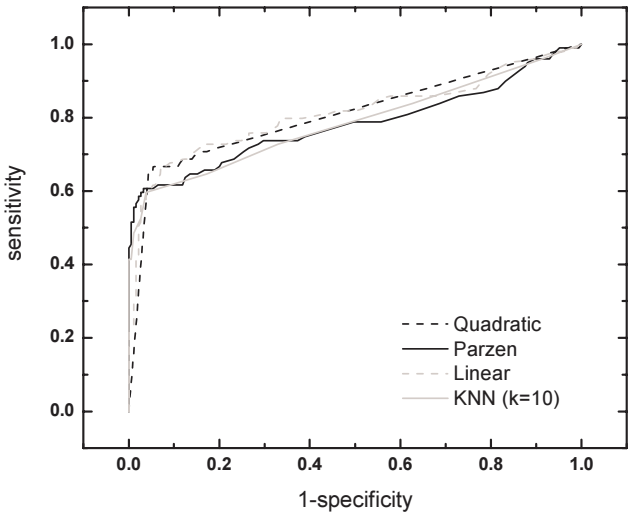


**Fig. 6**. ROC curves of the $k$NN ($k = 10$), Parzen, linear discriminant, and quadratic discriminant classifier. Az-values for these curves are 0.78, 0.78, 0.81, and 0.81 respectively. It can be seen that for a specificity of 0.95, the sensitivity can be higher than 0.65 (quadratic discriminant classifier).

classifier with $k = 10$, a Parzen classifier ($\sigma = 1.0$), a linear discriminant classifier and a quadratic discriminant classifier. A 10-fold cross-validation was used to asses the performance of the classifiers. Fig.6 depicts the ROC curves calculated for all four classifiers. For our specific problem, it is very important that the false positive rate is as low as possible, as the cost of misclassifying a potentially malignant nodule as a benign lymph node is very high. We see that sensitivities up to 65 percent can be reached at a specificity of 95 percent. A closer inspection of misclassified PFO's reveals that most of these nodules do not have a fissure in the neighborhood that can clearly be detected, even by eye. Without a detected fissure to relate to the flat sides of a nodule, performance of our algorithm is lower.

## 4. CONCLUSIONS

A new CAD scheme has been developed to distinguish between PFO's and other types of nodules generally found on high resolution CT images. Local shape features have been developed that take into account the predominantly flattened shapes that are characteristic for benign PFO's. Furthermore, a method has been developed for detecting fissures around the nodule. By calculating the similarity between the direction of the found fissure-planes around the nodule and the direction of planes fitted through the nodule's surface voxels, a high response to PFO's is obtained. The results suggest that our novel CAD scheme can reliably separate PFO's from other, potentially malignant, lesions. Because PFO's occur frequently, their exclusion could significantly reduce the false-positive rate in systems that aim to distinguish malignant nodules from benign ones.

## 5. REFERENCES

[1] Bethany B. Tan, Kevin R. Flaherty, Ella A. Kazerooni, and Mark D. Iannettoni, "The solitary pulmonary nodule," *Chest*, vol. 123, no. 90010, pp. 89S–96S, 2003.

[2] M.I. Ahn, A. McWilliams, S. MacDonald, S. Lam, and J. Mayo, "Perifissural Opacities (PFO's) Detected at Chest CT Screening for Lung Cancer," in *RSNA*, 2004, pp. 324–324.

[3] R. Wiemker, P. Rogalla, T. Blaffert, D. Sifri, O. Hay, E. Shah, R. Truyen, and T. Fleiter, "Aspects of computer-aided detection (CAD) and volumetry of pulmonary nodules using multislice CT," *British Journal of Radiology*, vol. 78, pp. 46–56, 2005.

[4] R. Wiemker, T. Bülow, and T. Blaffert, "Unsupervised extraction of the pulmonary interlobar fissures from high resolution thoracic CT data," *Computer aided radiology and surgery, CARS*, p. 1121, 2005.

[5] J. Staal, *Segmentation of elongated structures in medical images*, Ph.D. thesis, Universiteit Utrecht, the Netherlands, 2004.

[6] P.V.C Hough, "Machine Analysis of Bubble Chamber Pictures," *International Conference on High Energy Accelerators and Instrumentation*, 1959.