Supervised Enhancement Filters: Application to Fissure Detection in Chest CT Scans

Eva M. van Rikxoort*, Bram van Ginneken, *Member, IEEE*, Mark Klik, and Mathias Prokop

***Abstract*—In medical image processing, many filters have been developed to enhance certain structures in 3-D data. In this paper, we propose to use pattern recognition techniques to design more optimal filters. The essential difference with previous approaches is that we provide a system with examples of what it should enhance and suppress. This training data is used to construct a classifier that determines the probability that a voxel in an unseen image belongs to the target structure(s). The output of a rich set of basis filters serves as input to the classifier. In a feature selection process, this set is reduced to a compact, efficient subset. We show that the output of the system can be reused to extract new features, using the same filters, that can be processed by a new classifier. Such a multistage approach further improves performance. While the approach is generally applicable, in this work the focus is on enhancing pulmonary fissures in 3-D computed tomography (CT) chest scans. A supervised fissure enhancement filter is evaluated on two data sets, one of scans with a normal clinical dose and one of ultra-low dose scans. Results are compared with those of a recently proposed conventional fissure enhancement filter. It is demonstrated that both methods are able to enhance fissures, but the supervised approach shows better performance; the areas under the receiver operating characteristic (ROC) curve are 0.98 versus 0.90, for the normal dose data and 0.97 versus 0.87 for the ultra low dose data, respectively.**

***Index Terms*—Classifier, enhancement, Hessian matrix, pulmonary fissures, supervised.**

I. Introduction

STRUCTURE enhancement in volumetric data is an essential first step in many image analysis systems. To enhance structures in 3-D computed tomography (CT) data, many researchers have employed filters based on eigenvalues of the Hessian matrix or the structure tensor. This was originated by Haussecker *et al.* [1] who described the use of the structure tensor for local structure analysis and Koller *et al.* [2], who defined filters for the detection of curvilinear structures in 2-D and 3-D data based on the Hessian matrix. Frangi *et al.* [3] further developed this approach by defining a multi-scale tubular structure enhancement filter based on the eigenvalues of the Hessian matrix and applying it to the enhancement of vessels. Sato *et al.* [4] and Li *et al.* [5] defined enhancement filters for tubes, plates and blobs based on the eigenvalue analysis of the Hessian matrix. Both Sato *et al.* [4] and Li *et al.* [5] applied their filters to different medical data sets to show its efficiency. More recently, Agam *et al.* [6] defined a vessel enhancement filter based on the eigenvalues of the structure tensor and showed that it is less likely to respond to nodules compared to Hessian based filters. Most of the enhancement filters proposed so far, are effective in enhancing a certain structure, but not in all cases (e.g., vessel enhancement filters typically show a low response on bifurcations) and they tend to enhance other structures as well (e.g., a vessel filter may yield a higher response on a strong edge than on a low contrast tubular structure).

To overcome these problems, we propose supervised filters, that is, filters that are constructed using example input and output data and classifiers from pattern recognition theory [7]. Supervised methods are flexible: they can be used for many different structures provided the appropriate examples are used for training. In this work we will propose a multiphase process which optimizes the enhancement results by using not only features from the original scan but also from results from the previous phase(s).

A general problem of enhancement filters for volumetric data is validation of the results since setting a reference standard is time consuming and performing a quantitative evaluation that considers both sensitivity and specificity is difficult. For example, Sato *et al.* [4] evaluated the contrast between the target object and the background and the contrast to noise ratio for two sets of simulated data. They also applied their filters to four medical applications but no quantitative evaluation was provided. Li *et al.* [5] applied their filters to synthetic 2-D data, 2-D thoracic CT images and 3-D thoracic CT images. For both the synthetic 2-D data and the 3-D medical data no quantitative evaluation was provided. For the 2-D medical data the percentage of detected nodules together with the average false positive rate was provided. Agam *et al.* [6] do provide a quantitative evaluation for 600 CT slices. The mean square error between a vessel segmentation produced using an adaptive threshold segmentation and a thresholded outcome of their enhancement filter was provided. However, this performance evaluation does not take both sensitivity and specificity into account.

In this paper, we will evaluate the supervised approach for enhancing pulmonary fissures and compare the results to the results of a recently published "conventional" approach. The results will be evaluated on two different data sets, one normal dose data set containing 25 CT chest scans and an ultra-low dose data set containing 20 CT chest scans. The results will be quantitatively evaluated using receiver operating characteristic (ROC) analysis. In Section II, background and previous work regarding fissure detection will be provided. Section III describes the materials. Next, in Section IV both the supervised and nonsuper-

Manuscript received December 22, 2006; revised April 26, 2007. *Asterisk indicates corresponding author.*

*E. van Rikxoort is with the Image Sciences Institute, 3584CX, Utrecht, The Netherlands (e-mail: eva@isi.uu.nl).

B. van Ginneken and M. Klik are with the Image Sciences Institute, 3584CX, Utrecht, The Netherlands (e-mail: mark@isi.uu.nl; bram@isi.uu.nl).

M. Prokop is with the Department of Radiology, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands (e-mail: m.prokop@azu.nl).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2007.900447

vised filters will be described in detail. In Section V, a quantitative analysis of the results will be given, and finally in Section VI we discuss the results and draw conclusions.

II. Fissure Detection: Background and Previous Work

There are two types of pulmonary fissures, lobar fissures and accessory fissures. The lobar fissures divide the human lungs into five regions called lobes; the left lung consists of two lobes, the right lung of three lobes. Physically, interlobar fissures consist of a double layer of visceral pleura and constitute anatomical barriers to invasion of neoplastic and inflammatory diseases [8]. The lobar fissures are anatomic landmarks in the interpretation of CT scans; radiologic identification of a lesion in relation to the fissures is important for precise localization of the lesion with respect to the anatomic pulmonary lobes [9]. The lobar fissures may be incomplete, in which case the different lobes of the lung are partly connected.

Next to lobar fissures, there are accessory fissures. Physically, an accessory fissure is a cleft of varying depth lined by visceral pleura [10]. Accessory fissures often occur between bronchopulmonary segments but may also enter subsegmental or interbronchial planes [11]. Accessory fissures can be complete or incomplete and extend inward towards the hilum at different depths. Identifying accessory fissures is important for a more precise localization of lesions and characterization of diseases [12].

Recognition of pulmonary fissures is very hard on thick (10 mm) section chest CT scans, since due to the partial volume effect they are only visible as vague bands of increased density. Modern scanners provide thin section CT data in which the partial volume effect is substantially decreased and pulmonary fissures appear as bright plate-like structures.

Several methods have been published to automatically or semi-automatically segment the interlobar fissures and lung lobes on CT chest scans. Zhang *et al.* [13] used an anatomic lung atlas to initialize lobar fissure search. Next, a ridgeness measure was applied to 2-D slices to enhance fissure contrast and fuzzy logic was used to extract the final lobar fissure positions. They showed that their method can find the oblique lobar fissures with good precision. Kuhnigk *et al.* [14] detected the lobar fissures using a distance transform to a vessel segmentation and the original CT values. Next, an interactive 3-D watershed transform was used to segment the lobes. They showed that the final result of their lobe-segmentation did not vary substantially for different manual initializations. More recently, Ukil *et al.* [15] developed a method in which the search area for the major and minor fissures is first defined by the anatomy of the vessel and airway tree. Next, in the volume of interest, an optimal 3-D surface is found. Their method showed good results in segmenting the lobes. All these methods focus only on the interlobar fissures and are designed to segment the lobes. Our goal is not to segment the lobes, but to enhance all fissures present in the lungs.

This task was also addressed by Wiemker *et al.* [16] who applied a plate-enhancement filter to enhance pulmonary fissures. They developed two filters, one based on the eigenvalues of the structure tensor and one based on the eigenvalues of the Hessian matrix. In their paper, no quantitative evaluation of the results was given, but they concluded that both filters gave similar results. We implemented both filters and came to the same conclusion, therefore, in this paper the supervised approach will be compared to only one of these filters, namely the Hessian based filter. It should be noted that this Hessian based fissure enhancement filter is very similar to the plate enhancement filter proposed by Li *et al.* [5]. The main difference is that in the fissure enhancement filter some knowledge about the density of fissures on 3-D CT data is added. Details about the Wiemker filter are provided in Section IV-A.

TABLE I
Characteristics of the Data of the 25 Patients; Note That One Patient Can Have Multiple Pathologies

pathology	# appearances in patients
no pathology	8
ground glass	3
emphysema	7
bronchitis/bronchiolitis	3
interstitial fibrosis	1
fibrosis	2
consolidation	1
lung distortion	1
nodules	1

III. Data

For this study, 25 normal dose (120 kV, 100–150 mAs, volume CT dose index $= 7.8 - 11.6$ mGy, depending on patient size and weight) inspiration CT chest scans of 25 different patients were randomly selected from clinical practice at the University Medical Center Utrecht. For a subset of 10 of those patients, also an ultra-low dose (90 kV, 15–20 mAs, volume CT dose index $= 0.5 - 0.7$ mGy) inspiration and expiration CT chest scan were selected, making for a total of 45 CT chest scans. Data was acquired on an Mx8000 IDT CT-scanner or Brilliance 16P from Philips Medical Systems (Cleveland, OH) in about 12 s in spiral mode with 16×0.75 mm collimation and 15 mm table feed per rotation (pitch $= 1.3$). Axial images of 1.0 mm thickness at 0.7 mm increment were reconstructed with a moderately soft kernel (Philips "B") using the smallest field of view that included the outer rib margins at the widest dimension of the thorax. All scans were reconstructed with a 512×512 matrix, yielding an axial resolution in between 0.6 and 0.8 mm. In Table I, the characteristics of the patient data are described. Next to the pathologies present in the data, 10 scans contained breathing artifacts.

In all scans, the lung fields were segmented with an automatic 3-D algorithm described by Sluimer *et al.* [17], comparable to the algorithm proposed by Hu *et al.* [18], using standard image processing techniques such a region growing and morphological smoothing. In all cases, the lung segmentation was successful.

The data was divided into two sets, one normal dose set and one ultra-low dose set. Next, each dataset was divided into a training set and a test set. For the normal dose data, the training set contained 13 scans and the test set 12 scans, for the ultra-low dose data both sets contained 10 scans, of which five are inspiration scans and the other five were expiration scans of the same patients.

For the scans in the training and test sets, a human observer manually indicated the lobar fissures in every fourth coronal

slice. For the test sets, a second human observer performed the same task in order to be able to compare the performance of the filters with an independent human observer. Since fissures are often hard to distinguish using only 2-D information, before segmenting the fissures on the coronal slices, observers scrolled through the scan in the axial and sagittal direction and indicated marker points on each individual fissure. These points were visible in the coronal direction while segmenting the fissures. Segmenting fissures was performed by clicking points on the fissure; between two points, a straight line was automatically drawn. This line was refined using the Hounsfield values in a small neighborhood perpendicular to the line. Each voxel on the straight line between two manually indicated points was moved at most one voxel in the perpendicular direction to a voxel with higher Hounsfield values. This refinement was necessary to accurately segment the centerlines of the fissures. Examples of the data used and the manual segmentations can be seen in Fig. 3.

For the normal dose data, observers were also instructed to segment all accessory fissures in the scans. The accessory fissures were not manually segmented in the ultra-low dose scans since accessory fissures are practically invisible in these noisy images.

IV. Method

In this section, supervised and nonsupervised enhancement will be described and the particular choices for the application of fissure enhancement will be explained. For supervised enhancement two approaches will be described, a single phase method and a multiphase method.

A. Nonsupervised Enhancement

Conventional, nonsupervised enhancement filters are typically based on first (structure tensor) or second (Hessian matrix) order image information. These filters exploit the characteristics of idealized shapes like tubes and plates, sometimes extended with task-specific measures like intensity.

An enhancement filter based on second order information is based on gray-scale curvature information. To determine these curvatures, the eigenvalues of the Hessian matrix are used. For each voxel, the Hessian matrix is build from the six independent second order derivatives

$$H = \begin{bmatrix} L_{xx} & L_{xy} & L_{xz} \\ L_{yx} & L_{yy} & L_{yz} \\ L_{zx} & L_{zy} & L_{zz} \end{bmatrix}. \tag{1}$$

When performing an eigenvalue analysis of the Hessian matrix, the principal directions in which the local second-order structure can be decomposed are extracted. To construct a filter using this information, curvature characteristics of the structure to be enhanced are inspected. For example, in a tube one vanishing curvature is expected parallel to the tube and two strong curvatures are expected perpendicular to the tube. A filter can now be constructed based on this information.

As a nonsupervised fissure enhancement method we implemented the Hessian based fissure enhancement filter as described by Wiemker *et al.* [16]. The filter is based on the fact that in a plate like structure like a fissure, one strong gray-level curvature is expected perpendicular to the plate and two small curvatures are expected parallel to the plate. Therefore, a fixed combination of the two largest eigenvalues is used to determine a planeness value P which becomes 1 when the largest eigenvalue is significantly larger that the other eigenvalues and 0 if the two largest eigenvalues are the same. Since we are looking for bright plates, the largest eigenvalue is demanded to be negative

$$P = \frac{|\lambda_0| - |\lambda_1|}{|\lambda_0| + |\lambda_1|}, \text{ for } \lambda_0 < 0 \tag{2}$$

$$P = 0, \text{ otherwise} \tag{3}$$

with $|\lambda_0| \geq |\lambda_1| \geq |\lambda_2|$.

Next to this planeness value, some fissure-specific knowledge is added to the enhancement filter. Since Hounsfield values are physical measurements, fissures are expected always to have Hounsfield values within a certain range. Therefore, the planeness value P is multiplied by a Gaussian function which describes how well the Hounsfield value I at the filter center corresponds to the typical Hounsfield value μ for fissures, with a standard deviation σ. So, the final fissure filter then becomes

$$F = \exp\left(\frac{I - \mu^2}{2\sigma^2}\right) \cdot P. \tag{4}$$

In order to calculate the eigenvalues of the Hessian matrix, derivatives of the image are taken using Gaussian convolution. This process makes the non-supervised fissure filter scale-dependent. Following Frangi *et al.* [3] and Li *et al.* [5] we implemented a multiscale version of the nonsupervised fissure enhancement filter. In the final output of the multiscale enhancement filter, the maximum output over all scales is assigned to each voxel. After experimenting on the training data, we found that using only one scale, $\sigma = 1$ voxel, gave optimal performance.

B. Supervised Enhancement

In supervised enhancement, two stages can be distinguished, a training stage in which the system is developed, and a test stage in which the system is applied to previously unseen data. In the training stage, a number of voxels are sampled from the training images, a set of features is calculated for each voxel and a classifier is trained. In order to be able to train the classifier, a ground truth is required which gives for each voxel the class label (i.e., belonging to the class to be enhanced or not). In this study, this ground truth is provided by a human observer, as outlined in Section III.

For each voxel used for training the system, features are extracted that describe certain characteristics of that voxel. Since it is hard to predict which characteristics are most important for a certain structure, it is important to provide a rich set of features. In this paper, this set contains the output of Gaussian filters up to and including second order derivatives (L, L_x, L_y, L_z, L_{xx}, L_{xy}, L_{xz}, L_{yy}, L_{yz}, L_{zz}). These filters (except L) are sensitive to the orientation of structures in the image. The use of Gaussian derivatives to locally describe image structure is motivated by the Taylor expansion [19], which is a projection of a

signal on its derivatives. Since orientation is not the only important property of structures, it is important to have rotationally invariant features as well. Therefore, we added the gradient (L_i), which responds to edges, and the eigenvalues of the Hessian matrix $(|\lambda_0| \geq |\lambda_1| \geq |\lambda_2|)$ which denote the principal directions of the local curvature. Finally, we added the identity filter (L_0) which, together with (L), adds gray-value information to the feature set. All features, except L_0, can be calculated at different scales (σ), in this paper we use four scales, i.e., $\sigma = 1$, 2, 4, 8. In total we thus use 57 features in this paper; 40 features resulting from the Gaussian filters on four scales, 12 features come from the eigenvalues of the Hessian matrix, the gradient on four scales and L_0.

There is no reason to assume that all features as described above are important features for the structure to be enhanced. Therefore, feature selection can be applied in the training stage of the system. The main goal of feature selection is to select a subset of features for which the performance of the system does not substantially degrade. When only a subset of the original features is used, the test stage will be faster since less features have to be calculated and the dimensionality of the data is reduced.

There are many ways to perform feature selection. In this paper we use sequential floating forward selection (SFFS) as described by Pudil *et al.* [20]. SFFS is a wrapper-based algorithm, which means that it tests the classification performance of a specific classifier using different feature sets. Compared to other methods, this method has shown good performance for practical problems [21]. To perform SFFS the training data is divided into two sets, one set for training the classifier with a particular set of features and the other set for testing the performance of the classifier trained with those features. SFFS starts with an empty feature set. Next, in each iteration either a new feature is added or an already selected feature is removed depending on which gives the best performance on the optimization set, and in such a way that already examined configurations are not tested repeatedly. The final result of the feature selection will be the subset of features which performed best on the optimization set.

To apply the SFFS, a classifier must be selected. Many classifiers are described in literature. For this paper, four classifiers were examined, the linear discriminant classifier (LDC), the quadratic discriminant classifier (QDC), the support vector machine (SVM) with a Gaussian kernel, and the k-nearest neighbor classifier (knn) [7]. So, the outcome of the feature selection will be the optimal set of features for each specific classifier. The classifier–feature set combination that performs best was selected to be used in the test phase. The knn classifier outperformed the other classifiers with an area under the ROC curve (A_z) of 0.95 versus $A_z = 0.85$ for LDC, $A_z = 0.83$ for QDC, and $A_z = 0.93$ for SVM. For the knn classifier, the value of k was varied between 1 and 45 and found to be optimal at 15. Therefore, in the remaining of this paper, the knn classifier with $k = 15$ will be used (denoted as a 15 nn classifier).

C. Single Phase Supervised Filtering

With the prerequisites as described above, a single phase supervised filter can be defined. For training the system, voxels need to be sampled from both the positive and negative class. Next, features are calculated for these voxels, feature selection is applied and a classifier is trained and used to enhance the unseen test data.

For the purpose of fissure enhancement, we designed a single phase supervised filter with the following specifications: since only approximately 1% of the voxels are on a fissure, sampling voxels using a grid or random sampling does not lead to optimal performance. Therefore, as positive examples, all voxels indicated by the human observer as being fissure voxels were used. An equal amount of randomly selected voxels elsewhere in the lungs were selected as negative examples. Half of these negative examples were forced to be within five voxels of the positive examples, the other half was more than five voxels away from the positive examples. The 57 features as described in Section IV-B are used. The training data is divided into two sets of equal size to perform SFFS. The feature selection was set to optimize area under the ROC curve for the optimization set using the 15 nn classifier. With the features resulting from the feature selection procedure and the labels following from the manual segmentations, a 15 nn classifier was trained.

D. Multiphase Supervised Filtering

Since the single-phase supervised enhancement as described above uses only local information for each voxel, there are often spurious responses originating from the properties of a single voxel. To increase enhancement performance and suppress spurious responses, we developed a multiphase approach. This multiphase approach starts with a single-phase supervised filter, which results in a first output image (see third column of Fig. 1). This output image contains for each voxel a probability that it belongs to the structure to be enhanced. Next, another supervised filter is constructed extracting features from both the original image and the output image from the previous phase(s). All these features together are used to train a new classifier. This process can be iterated many times. The rationale behind this approach is that the structure to be enhanced will be more clear in the (intermediate) enhanced image than in the original image. So, adding features from these intermediate images will result in potentially more discriminative features between the voxels to be enhanced and the spurious responses from the previous phase(s).

The multiphase approach does not only improve performance but can also be used to speed up supervised enhancement. A simple classifier (e.g., using very few features) can be used in the first phase(s) to discriminate between "easy" voxels and more difficult voxels. Two thresholds can be set on the result, a lower and an upper threshold, which separates the easy voxels from the more difficult voxels, and only the difficult voxels are considered in the next phase(s). This thresholding approach might also improve results since the features selected in the feature selection procedure, together with the classifier, will be specialized in discriminating between the more difficult voxels.

Since the identity filter (L_0) is one of the features and feature selection is applied, it is expected that the results of the multiphase approach are unlikely to be worse than those of the single phase approach: If the previous result is optimal, the expected outcome of the feature selection would be that only the identity filter is selected.

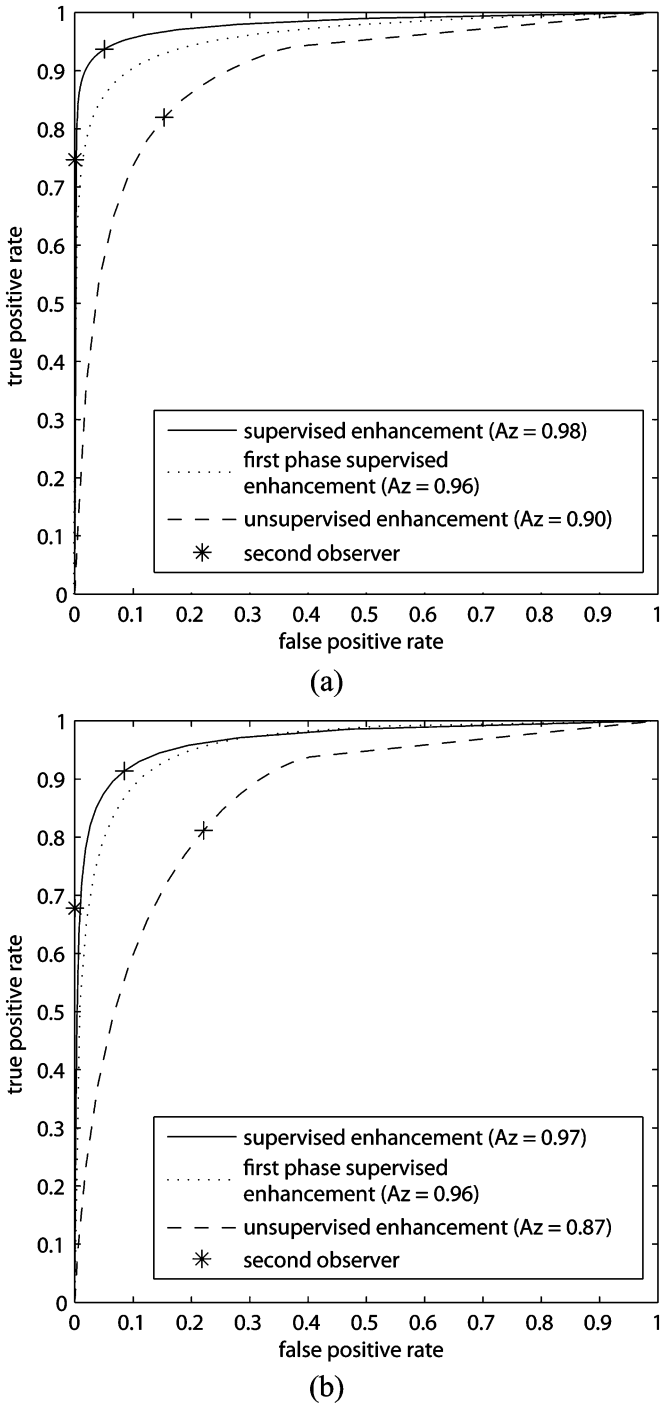


Fig. 1. (a) ROC curve for the 12 normal dose scans in the test set. A_z for the two-phase supervised method is 0.98, A_z for the single phase supervised enhancement is 0.96, A_z for the nonsupervised filter is 0.90. Pluses indicate the optimal threshold. Point of the second observer is also shown. (b) ROC curve for the 10 ultra low dose scans in the test set. A_z for the two-phase supervised method is 0.97, A_z for the single phase supervised method is 0.96, and A_z for the nonsupervised method is 0.87. Pluses indicate the optimal threshold. Point of the second observer is also shown.

For the application of fissure enhancement, a multiphase enhancement system was designed. After experimenting on the training data, a two-phase approach was chosen since using more phases did not substantially improve performance. The first phase is identical to the single-phase approach as described above. After this phase, a lower threshold is set at 0.2 to eliminate all voxels that are very probable to be nonfissure voxels. In the next phase, features are extracted from both the original data and the result from the first phase, which results in 114 features. Except for the additional features, the design of the second phase is identical to the design of the first phase.

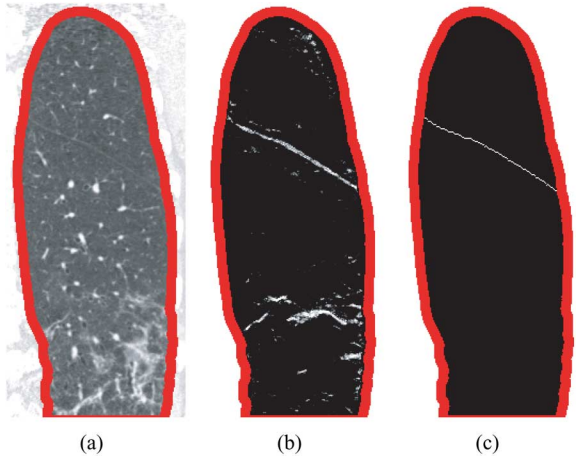


Fig. 2. Example output of the supervised method where false responses are given on pathology present in the data. Segmented lungs are indicated by the border around the lung volume. (a) Original slice. (b) Result of the supervised enhancement. (c) Manual segmentation of the fissure.

V. Experiments and Results

Two systems were developed, one for the normal dose data and one for the ultra low dose data. The results of both systems will be given separately. The systems were evaluated using ROC curves which capture the trade-off between sensitivity and specificity. Both the ROC curves and the area under the ROC curve were calculated using the trapezoidal rule. Since fissures were drawn on every fourth coronal slice, only these slices were used for evaluation, which results in 768 slices for the normal dose data and 649 slices for the ultra low dose data.

Although fissures are sometimes more than one voxel in width, the human observers were asked to draw a line with a width of one voxel in the center of the fissure. The reason for this is that the width of a fissure can vary even in the same slice and determining the exact boundary is not easy. To make sure the results of the systems were not influenced by this, a small band of five voxels around the manual segmentation is not taken into account to compute the ROC curve. As a consequence, voxels on the ground truth are positives, voxels close to the ground truth are not taken into account, and voxels more than five voxels away from the ground truth are regarded as negatives.

For the normal dose data, the supervised system was trained using the 13 scans in the training set for which one manual segmentation exists. Based on the feature selection, nine features were used in the first phase and 14 in the second phase. Table III lists those features. Due to the threshold of 0.2 that was set on the posterior probabilities of the first phase, only 35% of the voxels had to be reclassified in the second phase. Both the supervised and the nonsupervised method were applied to the 12 scans in the test set for which two manual segmentations exist to allow

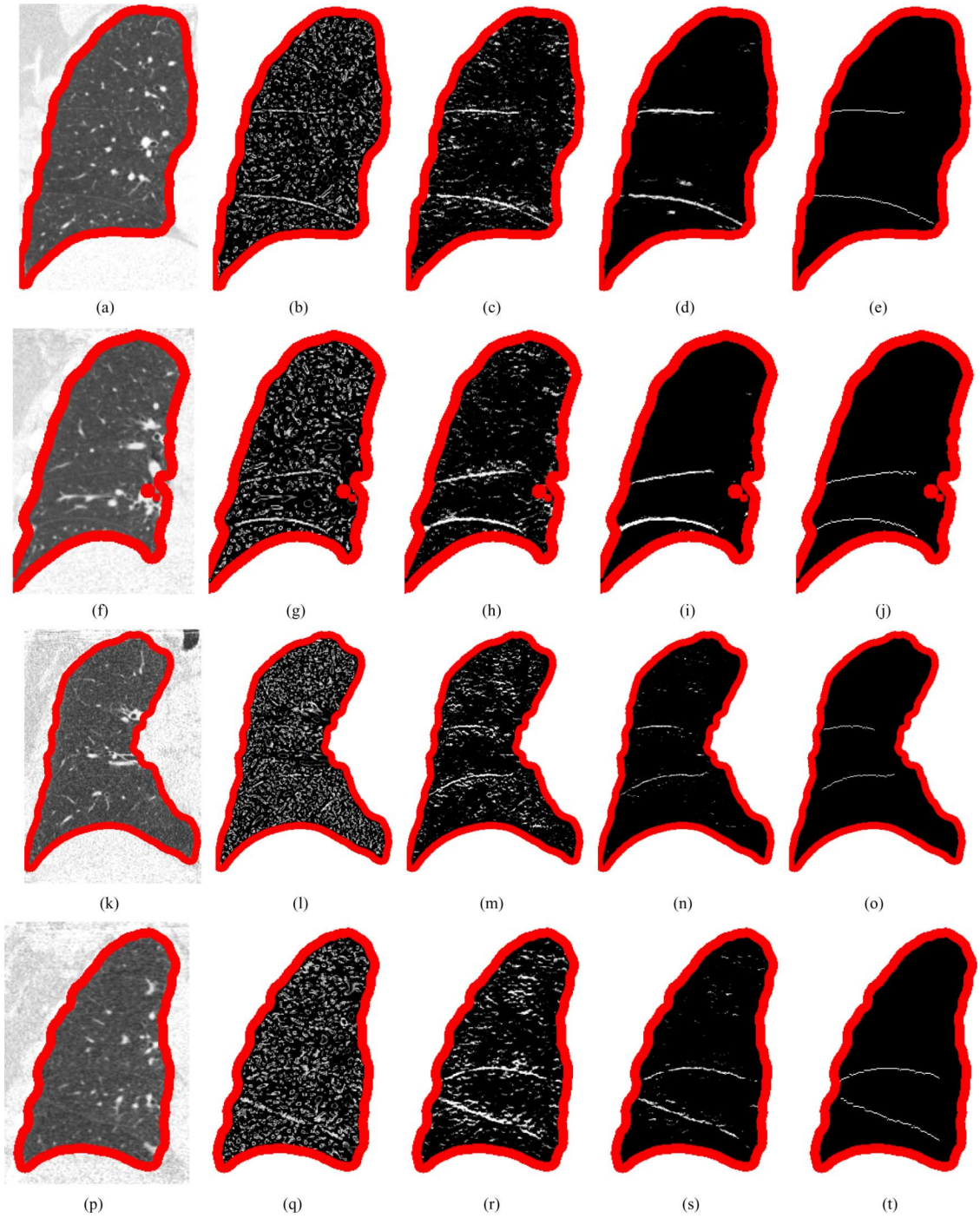
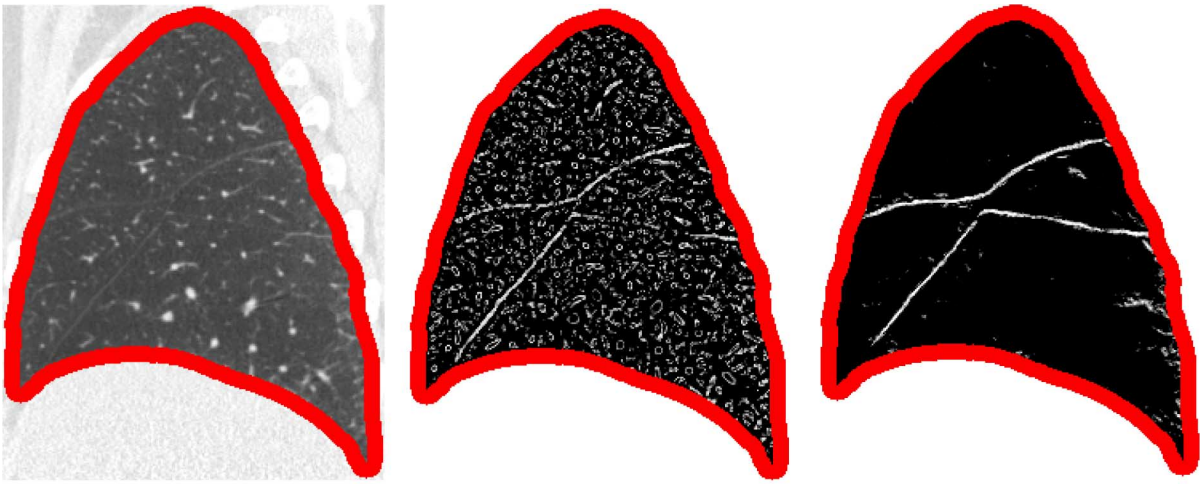


Fig. 3. Example output of both enhancement filters for a coronal slice of two normal dose scans (row 1 and 2), a coronal slice of an ultra-low dose inspiration scan (row 3), and an ultra-low dose expiration scan (row 4). Segmented lungs are indicated by the border around the lung volume. (a), (f), (k), (p) Original slice. (b), (g), (l), (q) Result of the nonsupervised enhancement. (c), (h), (m), (r) Results of the first step of the supervised enhancement. (d), (i), (n), (s) Final result of the supervised enhancement. (e), (j), (o), (t) Ground truth.

for quantitative evaluation. In the first two rows of Fig. 3 illustrative results are shown. In Fig. 4, the result of both methods are shown for a sagittal slice of the right lung on which the accessory segment 6 fissure is clearly visible. In Fig. 2, a coronal slice is shown where the supervised method produces false responses to pathology that is present in the data. Note that not all pathology leads to a response, but merely parts of it that resemble plate like structures. Both methods were quantitatively evaluated by means of ROC analysis, the results are shown in Fig. 1(a). In this plot also the performance of the second observer as compared to the first observer is shown. For the comparison of the second observer to the first observer a band of five voxels around the manual segmentation of the first observer is excluded, as was done for comparison with the automatic system. Based on the ROC curve an optimal threshold is defined as the point on the curve closest to (0,1), these points are indicated with plusses for both methods. In Table II, the areas under the curve for the separate fissures are given as well as the accuracy, sensitivity, and



(a) (b) (c)

Fig. 4. Example output of both enhancement filters for a sagittal slice of the right lung of a normal dose scan, appreciate the accessory segment six fissure. Segmented lungs are indicated by the border around the lung volume, the ground truth is not provided since it is only available for coronal slices. (a) Original slice. (b) Result of the nonsupervised enhancement. (c) Result of the supervised enhancement.

TABLE II
AREA UNDER THE CURVE FOR ALL FISSURES AND FOR THE SEPARATE FISSURES AS WELL AS ACCURACY, SENSITIVITY, AND SPECIFICITY FOR THE OPTIMAL THRESHOLD ARE GIVEN

	Normal dose	
	Supervised	**Non-supervised**
A_z total	0.98	0.90
A_z major	0.91	0.86
A_z minor	0.91	0.84
A_z accessories	0.91	0.86
	After optimal threshold	
accuracy	0.95	0.85
sensitivity	0.94	0.82
specificity	0.95	0.84
	Ultra low dose	
	Supervised	**Non-supervised**
A_z total	0.97	0.87
A_z major	0.97	0.87
A_z minor	0.97	0.86
A_z accessories	-	-
	After optimal threshold	
accuracy	0.92	0.78
sensitivity	0.91	0.81
specificity	0.92	0.78

TABLE III
OPTIMAL SET OF FEATURES SELECTED IN FEATURE SELECTION FOR EACH DATA SET, FOR THE SECOND PHASE ONLY FEATURES EXTRACTED FROM THE RESULT OF THE FIRST PHASE WERE SELECTED BY THE FEATURE SELECTION

Selected features			
Normal dose		**Ultra-low dose**	
first phase	**second phase**	**first phase**	**second phase**
$L_{zz},\sigma = 1$	$L_{yz},\sigma = 8$	$L_{zz},\sigma = 1$	$\lambda_0,\sigma = 1$
$\lambda_0,\sigma = 4$	$L_i,\sigma = 8$	$\lambda_0,\sigma = 4$	$\lambda_1,\sigma = 2$
$\lambda_1,\sigma = 2$	$L,\sigma = 8$	$\lambda_1,\sigma = 2$	$\lambda_0,\sigma = 2$
$L_i,\sigma = 8$	$\lambda_0,\sigma = 8$	$\lambda_0,\sigma = 2$	$L_{zz},\sigma = 1$
$L_i,\sigma = 1$	$L_{yz},\sigma = 2$	$L_i,\sigma = 1$	$\lambda_0,\sigma = 8$
$\lambda_0,\sigma = 2$	$L_{xx},\sigma = 4$	$L_{zz},\sigma = 2$	$L_{zz},\sigma = 1$
$\lambda_1,\sigma = 1$	$L_{yy},\sigma = 8$	$\lambda_0,\sigma = 8$	$L_z,\sigma = 1$
L_0	$L_{zz},\sigma = 4$	L_0	$L_{yy},\sigma = 1$
$L_{yz},\sigma = 1$	$L_{zz},\sigma = 1$	$L_{zz},\sigma = 4$	$L_{yz},\sigma = 2$
	$L_i,\sigma = 1$	$L_i,\sigma = 4$	$L_{yz},\sigma = 8$
	$L_{xx},\sigma = 8$	$\lambda_1,\sigma = 4$	L_0
	$L,\sigma = 4$	$L_{yy},\sigma = 4$	$L_i,\sigma = 8$
	$L_{yz},\sigma = 1$	$\lambda_0,\sigma = 1$	$L_i,\sigma = 1$
	$L_{xz},\sigma = 2$	$L_{zz},\sigma = 2$	$\lambda_0,\sigma = 4$
		$L_{zz},\sigma = 8$	
		$L_i,\sigma = 2$	

specificity at the optimal threshold. To be able to calculate the A_z values for the different fissures separately, the positions of the other fissures as given by the ground truth were not taken into account for those ROC analyses.

The supervised method for the ultra-low dose scans was trained on the 10 scans in the training set. Note that five of these scans are inspiration scans and the other five are expiration scans. Based on the feature selection 16 features were selected for the first phase and 14 for the second phase, see Table III. After thresholding the first step only 27% of the voxels were left for the second step. In the two bottom rows of Fig. 3, a result is shown for both an inspiration and an expiration scan. For the ultra-low dose data the same evaluation was performed as for the normal dose data, see Fig. 1(b) and Table II. Note that no A_z value is given for the accessory fissure on ultra-low dose data since they were not marked by the human observers.

Both for the normal dose data and for the ultra-low dose data we performed a two tailed paired *t*-test on the A_z values per scan to see if the results of the supervised and unsupervised method were significantly different. This test merely assumes that those values obtained from different scans are independent. The results show that for both the normal dose data and the ultra-low dose data the A_z values are significantly different $(p < 0.001)$.

As an additional experiment, we applied our normal dose system to data with thicker slices. To simulate increased slice thickness, slices of the 12 scans in the normal dose test data set were averaged in the axial direction. Fig. 5 shows a result for several slice thicknesses. Note that no new system was trained, we applied the system trained with the original normal dose training data. In Table IV, the results of these experiments are summarized.

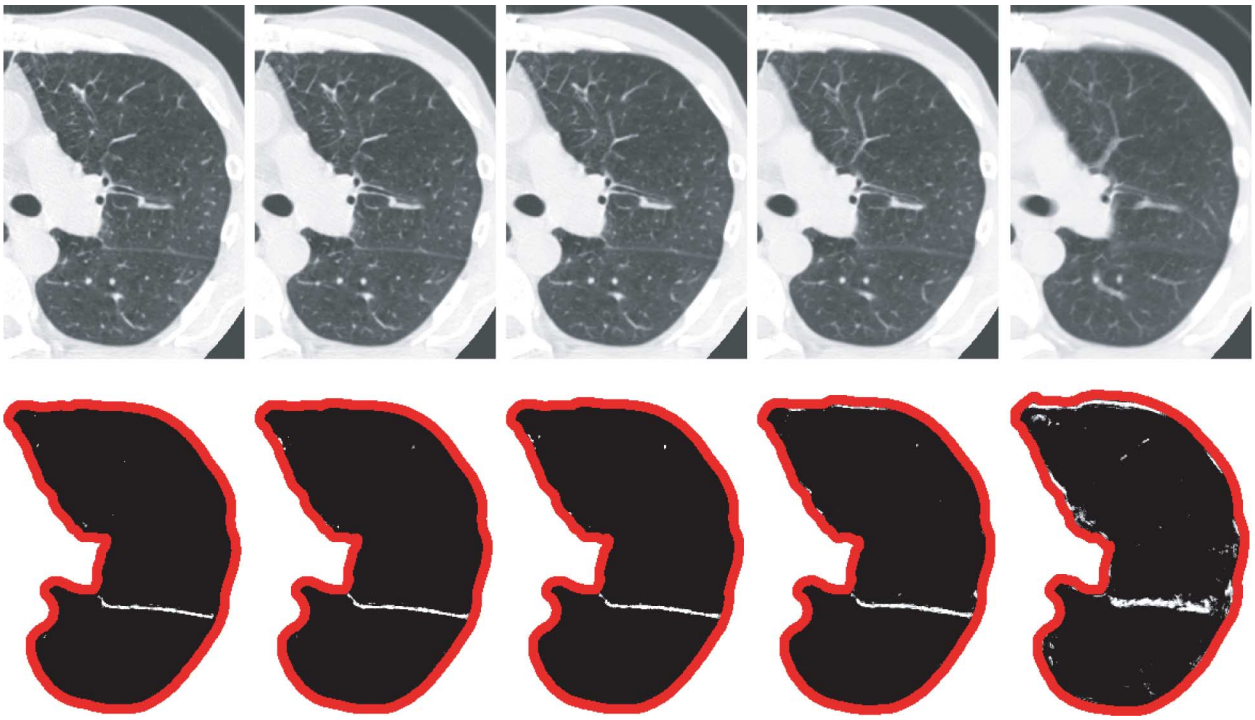


Fig. 5. Axial slice of the results of the supervised method for increasing slice thickness. Top row shows the original slice, the bottom row shows the corresponding result. In the first column, the original slice thickness of 1 mm is shown, moving to the right slice thickness is 1.7, 2.4, 3.8, and 7.3 mm, respectively.

TABLE IV
RESULTS OF AN ADDITIONAL EXPERIMENT IN WHICH LARGER SLICE THICKNESS IS SIMULATED IN THE NORMAL DOSE TEST DATA. IN THE FIRST COLUMN THE NUMBER OF AVERAGED SLICES IS SHOWN, THE SECOND COLUMN GIVES THE CORRESPONDING SIMULATED SLICE THICKNESS IN MILLIMETER, AND THE LAST COLUMN LISTS THE AREA UNDER THE ROC CURVE OF THE SUPERVISED FILTER

# slices	mm	A_z
1	1	0.98
2	1.7	0.98
3	2.4	0.97
5	3.8	0.92
10	7.3	0.69

VI. DISCUSSION AND CONCLUSION

In this paper, a pattern recognition approach to enhancing structures was presented and applied to enhancing pulmonary fissures. Although many attempts have been made to enhance structures in 3-D CT data using the eigenvalues of the Hessian matrix or the structure tensor, this paper is the first in which a pattern recognition approach is applied and compared to a recently published "conventional" filter approach. In addition, both filters were quantitatively evaluated on two large data sets. Both inspection of the images and an ROC analysis show that the supervised method gives superior results.

Next to better performance, there are other reasons to prefer a supervised approach to a nonsupervised approach. First of all, the supervised approach is flexible. It can be used for any structure provided the appropriate examples, whereas the nonsupervised approach requires the design of a completely new and different system for each specific structure. Secondly, supervised approaches usually are less sensitive to the settings of specific parameters than non-supervised approaches are, especially when using feature selection. In addition, conventional approaches can be interpreted as designed classifiers; the features and classification function are designed for the specific task. By including the same features in a supervised approach, a similar filter can be designed depending on the classifier used. So, when the features used in the nonsupervised approach are optimal, they will be selected during feature selection. Therefore, the supervised approach is likely never to perform worse than the conventional approach.

To further optimize the performance of supervised filters, a multiphase approach was introduced. In this multiphase approach, results from previous phases are used in each successive phase. This multiphase approach improves enhancement results substantially. The main improvement as opposed to a one-phase approach is the removal of spurious responses.

A drawback of a supervised approach is the fact that manual segmentations are required to train the system. Manually segmenting structures in 3-D CT data is labor intensive and when new data is used, e.g., from a different scanner or obtained with different acquisition parameters, the use of new training data might be required to obtain optimal performance. Although the use of new training data might improve the system it is not always necessary. This is illustrated by the experiment described in Section V where the normal dose fissure enhancement system is applied to simulated thicker slices without constructing a new system. This experiment shows that the supervised system is relatively insensitive to slice thickness and produces good results up to 3 mm slices. If data with thicker slices is processed with a filter trained with thin slice data, results deteriorate. This is not surprising and may be partly overcome by training with thick slice data as well, although one has to keep in mind that thin structures like fissures are easily blurred away in thick slice data and may not be visible even to human observers. As an

additional illustration we applied the normal dose supervised system to the ultra low dose data. This resulted in A_z of 0.93, which is lower than the performance of the supervised system trained with ultra-low dose data ($A_z = 0.97$) but still much better than the unsupervised method ($A_z = 0.87$).

Another possible drawback of supervised filters compared to conventional filters is longer computation time. The systems were implemented in C++, the code was not optimized. Results were calculated on a 3 GHz Pentium 4 machine. Applying the unsupervised fissure filter took on average 18 min. For the supervised filter, the computation time can be divided into an offline and online calculation time. The offline calculations consist of calculating the features, performing feature selection and training the classifier, and they need to be calculated once for every phase. The calculation of the features took on average 5 min per scan. The time for the feature selection depends heavily on the number of samples and classifier used. In case of the knn classifier, the feature selection took around 5 h for our system. Training the 15-nn classifier took 7 s. The online calculations for the fissure enhancement system consist of applying the trained classifier to the unseen test data. For the first phase this took on average 1 h, for the second phase this took on average 20 min. The time needed to calculate the supervised results could possibly be optimized by a different design of the multiphase approach; by using more phases and starting with fast, simple classifiers and processing only voxels with probabilities close to 0.5 in subsequent phases a large improvement in computation time could be achieved.

To be able to evaluate the supervised approach and compare it to a nonsupervised enhancement filter, a supervised fissure enhancement filter was developed. Both a nonsupervised and supervised filter were applied to three types of clinical data; normal dose CT chest scans, ultra-low inspiration CT chest scans and ultra-low expiration CT chest scans. The ultra-low dose inspiration and expiration scans were evaluated simultaneously since practically no differences were present between the results of these two types of data. Quantitative analysis of the results showed that the supervised approach is superior for all types of data. Inspection of the resulting images reveals the main difference between the two approaches; both filters are able to enhance the fissures but the nonsupervised approach also responds to many other structures (e.g., vessel walls) where the supervised approach has almost no false responses. When the supervised approach does produce a false response this is caused by structures that locally resemble a fissure. Inspecting the results showed that in our test data the false responses were mainly caused by fibrosis. An example is given in Fig. 2. Other pathology did not influence the performance of the supervised filter. Pathologies that change the appearance of the fissures or that locally resemble a fissure might pose a problem, in those cases new training data would be needed.

Both the results of the unsupervised and the supervised filters were compared to the performance of a second human observer. The output of the supervised filter is very similar to the output of the second observer, and this is not the case for the nonsupervised filter output. The relatively low sensitivity of the second observer as compared to the first observer can be explained by the fact that the first observer segmented more fissures in areas where these were barely visible. Observers were instructed to segment fissures where they could see them and this inevitably introduces some subjectivity. This indicates that segmenting the fissures is not an easy task for humans.

For the supervised method the performance on the ultra-low dose data shows no substantial difference both for the quantitative analysis and the inspection of the images as compared to the normal dose data. For the nonsupervised approach however there is a substantial difference. This is not only shown by the difference in area under the ROC curve ($A_z = 0.90$ versus $A_z = 0.87$) but also evident from inspection of the images. Although the non-supervised approach shows many false positive responses for the normal dose data, the fissure can usually be distinguished quite well. For the ultra-low dose data however this is not the case, in many slices the fissures are hardly distinguishable (see Fig. 3).

The effect of the multiphase approach is nicely illustrated by the fissure enhancement results. Although the ROC curves for the first phase are already substantially better than the ROC curves for the nonsupervised approach they are even further improved after the second phase. The difference between the two phases is also clearly seen on the resulting images (see Fig. 3). As can be seen in Table III different features were selected for both phases, which indicates that the second phase is specialized in discriminating between the more difficult samples. Interestingly, only features from the result of the first phase were selected for the second phase. This is probably due to the fact that the fissures are already enhanced quite well in the first phase, apart from spurious responses. Since these spurious responses are usually not bright plates, as fissures are, only the result of the first phase was necessary to remove these false responses.

In summary, a multiphase supervised approach to enhancement was presented and applied on the task of fissure enhancement on three types of 3-D CT chest scans. Evaluation of the results showed that the supervised approach is superior to a nonsupervised approach. In addition, the results of the supervised approach did not deteriorate substantially on much noisier data where the results of the nonsupervised approach did.

REFERENCES

[1] H. Haussecker and B. Jähne, "A tensor approach for local structure analysis in multi-dimensional images, in 3-D," in *Image Anal. Synthesis*, 1996, pp. 171–178.

[2] T. Koller, G. Gerig, G. Szekely, and D. Dettwiler, "Multiscale detection of curvilinear structures in 2-D and 3-D image data," in *Int. Conf. Computer Vision*, 1995, pp. 864–869.

[3] A. Frangi, W. Niessen, K. Vincken, and M. Viergever, "Multiscale vessel enhancement filtering," in *Med. Image Computing Computer-Assisted Intervention*, 1998, vol. 1496, pp. 130–137.

[4] Y. Sato, C. Westin, A. Bhalerao, S. Nakajima, N. Shiraga, S. Tamura, and R. Kikinis, "Tissue classification based on 3-D local intensity structure for volume rendering," *IEEE Trans. Vis. Comput. Graphics*, vol. 6, no. 2, pp. 160–180, Apr./Jun. 2000.

[5] Q. Li, S. Sone, and K. Doi, "Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans," *Med. Phys.*, vol. 30, pp. 2040–2051, 2003.

[6] G. Agam, S. G. Armato, III, and C. Wu, "Vessel tree reconstruction in thoracic CT scans with application to nodule detection," *IEEE Trans. Med. Imag.*, vol. 24, no. 4, pp. 486–499, Apr. 2005.

[7] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

[8] M. Gulsun, O. Ariyurek, R. Comert, and N. Karabulut, "Variability of the pulmonary oblique fissures presented by high-resolution computed tomography," *Surgical Radiologic Anatomy*, vol. 7, pp. 1–7, 2006.

[9] K. Hayashi, A. Aziz, K. Ashizawa, H. Hayashi, K. Nagaoki, and H. Otsuji, "Radiographic and CT appearances of the major fissures," *Radiographics*, vol. 21, no. 4, pp. 861–874, 2001.

[10] O. Ariyurek, M. Gulsun, and F. Demirkazik, "Accessory fissures of the lung: Evaluation by high-resolution computed tomography," *Eur. Radiol.*, vol. 11, no. 12, pp. 2449–2453, 2001.

[11] A. Yildiz, F. Gölpinar, C. Mukadder, M. N. Ducea, C.Özer, and F. D. Apydin, "HRCT Evaluation of the accessory fissures of the lung," *Eur. J. Radiol.*, vol. 49, pp. 245–249, 2004.

[12] M. 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.

[13] L. Zhang, E. A. Hoffman, and J. M. Reinhardt, "Atlas-driven lung lobe segmentation in volumetric x-ray CT images," *IEEE Tran. Med. Imag.*, vol. 25, no. 1, pp. 1–16, Jan. 2006.

[14] J.-M. Kuhnigk, V. Dicken, S. Zidowitz, L. Bornemann, B. Kuemmerlen, S. Krass, H.-O. Peitgen, S. Yuval, H.-H. Jend, W. S. Rau, and T. Achenbach, "New tools for computer assistance in thoracic CT part 1. Functional analysis of lungs, lung lobes and bronchopulmonary segments," *Radiographics*, vol. 25, pp. 525–536, 2005.

[15] S. Ukil, M. Sonka, and J. M. Reinhardt, "Automatic segmentation of pulmonary fissures in x-ray CT images using anatomic guidance," *SPIE Med. Imag.*, vol. 6144, pp. 61440N1–61440N11, 2006.

[16] R. Wiemker, T. Bülow, and T. Blaffert, "Unsupervised extraction of the pulmonary interlobar fissures from high resolution thoracic CT data," in *Computer Assisted Radiol. Surgery*, Berlin, 2005, pp. 1121–1126.

[17] I. Sluimer, M. Prokop, and B. v. Ginneken, "Towards automated segmentation of the pathological lung in CT," *IEEE Trans. Med. Imag.*, vol. 24, no. 8, pp. 1025–1038, Aug. 2005.

[18] S. Hu, E. A. Hoffman, and J. M. Reinhardt, "Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images," *IEEE Trans. Med. Imag.*, vol. 20, no. 6, pp. 490–498, Jun. 2001.

[19] B. H. R. ter, *Front-End Vision and Multi-Scale Image Analysis*, 1st ed. Dordrecht, The Netherlands: Springer-Verlag, 2003.

[20] P. Pudil, J. Novovicova, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Lett.*, vol. 15, pp. 1119–1125, 1994.

[21] A. Jain and D. Zongker, "Feature selection: evaluation, application and small sample performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 153–158, Feb. 1997.