Automatic Detection of Pulmonary Embolism in CTA Images

Henri Bouma, Jeroen J. Sonnemans, Anna Vilanova and Frans A. Gerritsen

***Abstract*—Pulmonary embolism (PE) is a common life-threatening disorder for which an early diagnosis is desirable. We propose a new system for the automatic detection of PE in contrast-enhanced CT images. The system consists of candidate detection, feature computation and classification. Candidate detection focusses on the inclusion of PE – even complete occlusions – and the exclusion of false detections, such as tissue and parenchymal diseases. Feature computation does not only focus on the intensity, shape and size of an embolus, but also on locations and the shape of the pulmonary vascular tree. Several classifiers have been tested and the results show that the performance is optimized by using a bagged tree classifier with two features based on the shape of a blood vessel and the distance to the vessel boundary. The system was trained on 38 CT data sets. Evaluation on 19 other data sets showed that the system generalizes well. The sensitivity of our system on the evaluation data is 63% at 4.9 false positives per data set, which allowed the radiologist to improve the number of detected PE by 22%.**

***Index Terms*—Pulmonary embolism, computed tomography, computed-aided detection.**

I. Introduction

PULMONARY EMBOLISM is the sudden obstruction of a blood vessel in the lungs, usually due to a blood clot. There is more than one case of PE per 500 persons every year in the USA [1]. Of these cases, 11% die in the first hour [2] and the untreated mortality of PE is estimated at 30% [3]. Thus, PE is a common disorder with a high morbidity and mortality for which an early and precise diagnosis is desirable [4].

Contrast-enhanced multi-slice x-ray computed tomography (CT) is emerging as the preferred imaging test to diagnose PE in many institutions, because it is a fast, minimally invasive and high-resolution imaging technique [5]–[8] that allows the direct depiction of a clot inside arteries. CT images can also be used to identify other disorders in a patient with chest pain [9].

In contrast-enhanced CT (i.e., CT angiography, CTA) images, the blood vessels appear to be very bright because the contrast material is dissolved in blood. The embolus does not absorb this material, and thus it can be recognized in CTA as a dark area in the pulmonary arteries (see Fig. 1a). However, manual detection of the dark spots that correspond to PE in CTA images is often described by radiologists as difficult and time consuming [10]. Therefore, computer-aided diagnosis (CAD) is desirable.

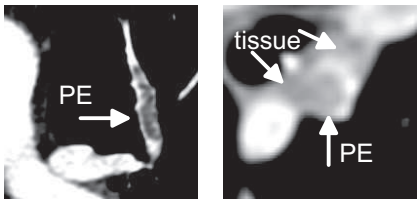


Fig. 1. (a) In CT images, PE appears as a dark region inside the pulmonary arteries. (b) Lymphoid tissue and PE can have the same intensity.

In the last years, several CAD systems for PE have been proposed, but their evaluations have some serious limitations. For example, for some system [11], [12] it is uncertain how well they generalize because a low number of emboli is used for evaluation. For another system [13], it is unclear how it performs on all emboli, because only the peripheral emboli are taken into account. And it is unknown how a system [14] performs on representative data when only data is used with good opacification and without significant motion artifacts or pulmonary diseases. (Thus it is not surprising that another evaluation of the same system with more realistic data showed much worse results [15].) The studies of Buhmann [16], Zhou [17] and Kiraly [18] are the only studies that reported the system performance on a large database of realistic test data that contained breathing artifacts and parenchymal diseases without the exclusion of emboli based on size or location. However, results of Buhmann and Zhou are poor when all emboli are taken into account (with sensitivities of 47% and 52% at 3.9 and 11.4 false positives per data set respectively). The 3D visualization method of Kiraly showed 50% sensitivity, which would improve the initial reading by 13%.

In this paper, we propose a new CAD system for automatic detection of PE in CTA images. For the training of the system, we used 38 positive data sets (202 PE), and for the evaluation, 19 other positive data sets (116 PE) were used. The evaluation shows that the performance of our system is at the level of state of the art literature. The data sets were selected to demonstrate considerable motion artifacts (15%), sub-optimal

Manuscript received July 25, 2008; revised December 06, 2008. First published February 10, 2009; current version published July 29, 2009. *Asterisk indicates corresponding author*.

*H. Bouma was with the TU Eindhoven, 5600 MB Eindhoven, The Netherlands. He is now with TNO, 2509 JG The Hague, The Netherlands. (e-mail: henri.bouma@tno.nl).

J.J. Sonnemans is with the Healthcare Informatics Division, Philips Healthcare, 5680 DA Best, The Netherlands (e-mail: jeroen.sonnemans@philips.com).

A. Vilanova is with the Department of Biomedical Engineering, TU Eindhoven, 5600 MB Eindhoven, The Netherlands (e-mail: a.vilanova@tue.nl).

F.A. Gerritsen is with the Healthcare Informatics Division, Philips Healthcare, 5680 DA Best, The Netherlands and also with TU Eindhoven, 5600 MB Eindhoven, The Netherlands (e-mail: frans.gerritsen@philips.com).

Digital Object Identifier 10.1109/TMI.2009.2013618

contrast (20%), a variety of thrombus load (25% with an obstruction degree above 25%) and parenchymal diseases (40%), and none of the emboli were excluded for evaluation. This is important because the main problem of PE detection is the separation between true PE and look-alikes, which is much harder when the patient has no other lung abnormalities.

The CAD system that we propose consists of several steps, which are described in three sections. In the first step, pulmonary vessels are segmented and PE candidates are detected inside the vessel segmentation (Sec. II). The candidate-detection step focusses on the inclusion of PE and the exclusion of false detections, such as lymphoid tissue and parenchymal diseases. Subsequently, features are computed on each candidate to enable classification of them (Sec. III). Feature computation does not only focus on the intensity, shape and size of an embolus, but also on the shape of the pulmonary vascular tree and the location in the tree. In the last step, classification is used to separate candidates that represent real emboli from the other. The system is optimized with feature selection and classifier selection (Sec. IV). The presented system for embolus detection is then evaluated and results are presented in Sec. V. Finally, the discussion and conclusions can be found in Sec. VI and VII respectively.

II. Vessel Segmentation and Candidate Detection

We propose a vessel-segmentation method to reduce the search area for candidate detection that includes (at least a part of) the emboli and it excludes (most of) the false detections and look-alikes. Examples of look-alikes that have the same intensity as PE are: lymphoid tissue (Fig. 1b), parenchymal diseases and partial-volume voxels on the vessel boundary. Another focus of our method is the inclusion of 'sudden stops' (i.e., completely embolized vessels), which are easily missed by common segmentation techniques. Figure 2 shows the design of our CAD system. The part about vessel segmentation and candidate detection is discussed in this section.

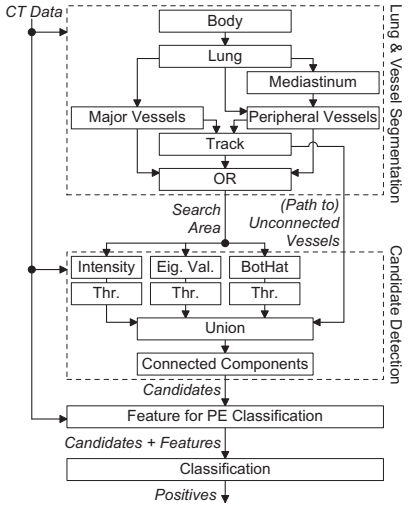


Fig. 2. Our CAD system for pulmonary embolism consists of vessel segmentation, candidate detection, feature computation and classification.

A. Vessel Segmentation

A rough selection of the lung region can be obtained in CT images by taking the two largest dark (below -500 Hounsfield Units, HU) regions in the thorax that have a minimum overlap of 30% when they are projected on the transverse axis [19]. These regions correspond to lung parenchyma (i.e., the cellular substance inside the lungs). To remove holes from the segmentation and include small vessels, a morphological closing [20] was used of 2.4 mm, which appeared to include most small vessels. A segmentation of the mediastinum (i.e., the region that contains the heart) was obtained by selecting the large region between the lungs using a row-wise operation that sets each pixel between the lungs. A morphological opening (6.0 mm) is applied to obtain a smooth mediastinal segmentation. To remove tissue from the lung segmentation, all voxels near the mediastinum segmentation (geodesic distance $<$ 11 mm in a mask with intensity $>$ -300 HU) are excluded.

The major pulmonary vessels can be segmented with a threshold, because they are brighter than other objects in the lungs. Region growing is performed to extend the segmentation to the mediastinum. This threshold is adapted to each data set, because of varying contrast enhancement. It is selected as the average of 0 HU and the most frequent intensity in the region with intensities above +150 HU.

The high threshold value used for major vessels cannot be used for the small pulmonary vessels. Their lower intensity is caused by the point-spread function (PSF), which makes it hard to create an intensity-based separation between vessels and (lymphoid) tissue. Small vessels are located in the periphery of the lungs and most of the tissue is near the mediastinum. The peripheral-vessel segmentation is selected as the region inside the lungs that is brighter than -150 HU to exclude lung parenchyma, with a radius below 2.4 mm to exclude large vessels, and with a distance to the mediastinum of at least 15 mm to exclude tissue near the mediastinum.

Tracking aims at the filling of a possible gap due to a complete obstruction between peripheral vessels and the rest of the pulmonary vascular tree. In healthy vessels without emboli, the segmentations of peripheral and major vessels are well connected (and overlapping). However in patients with PE, the obstruction can be so large that the segmentation of a peripheral vessel is 'unconnected' to other vessels. Locally it may be impossible to make a distinction between an embolized vessel and tissue because they have the same intensity. Therefore, we are using a tracking algorithm to find the optimal connection between the unconnected and other vessels. The tracking algorithm [21] uses three input segmentations. The first is the segmentation of the main tree, which includes all major and peripheral vessels that are connected to the mediastinum. The second is the segmentation of the unconnected vessels, which are all major and peripheral vessels that are *not* connected to the mediastinum. The third is a search area, which is inside and between the lungs with an intensity above -150 HU. The tracker is connecting the second region (unconnected vessels) to the first region (main tree) through the third region (search area). Wave-front propagation is used to compute costs for each voxel based on intensity and distance. Paths between

unconnected vessels and the main tree that have minimal costs are selected to create the connections. The dilated region around this path is combined with the segmentations of the major and peripheral vessels using an OR-operation to form the search area for candidate detection. The segmentations of the tracked path and the unconnected vessels are also used in candidate detection and feature computation.

B. Candidate Detection

Candidate detection aims at the extraction of a group of voxels inside the pulmonary vessels that includes (at least partially) an embolus. It reduces the search area by detecting voxels and grouping connected voxels to candidate objects. The area reduction allows an efficient feature computation and the grouping allows classification of these objects.

Four methods are used to detect candidate voxels inside the vessel segmentation. One of them is the union of the two binary masks found by tracking (Sec. II-A); the mask of the unconnected vessels, and that of the dilated path between these vessels and the main tree. Both masks are combined with the masks of the other detection methods using an OR-operation.

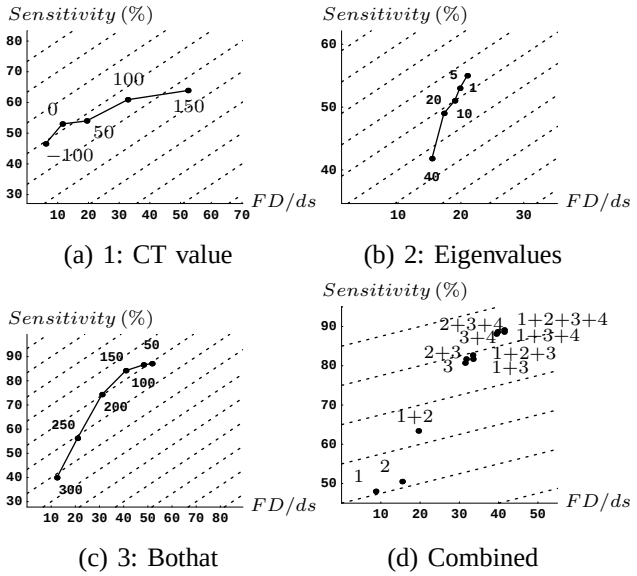
The other three methods are intensity-based features. The first directly uses the CT value. Inside the vessels of CTA images, emboli are darker than contrast-enhanced blood. The CT value allows a separation between these regions. The second uses the eigenvalues of the Hessian matrix (i.e., matrix of second-order derivatives) [11]. A dark spot can be detected with a positive first eigenvalue λ_1 of the Hessian, assuming that the eigenvalues are sorted by decreasing magnitude ($|\lambda_1| > |\lambda_2| > |\lambda_3|$). We used Gaussian derivatives at σ = 0.9 mm, since it is the lowest scale to take reliable derivatives on our data with voxels of 0.6 mm in each direction.

$$C_{\lambda_1} = \begin{cases} \lambda_1, & \lambda_1 > 0 \\ \lambda_3, & \lambda_1 \leq 0 \text{ and } \lambda_2 > 0 \text{ and } \lambda_3 > 0 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

The third is the (grey scale) bothat (black tophat [22] or local contrast [11]) with 4 mm dilations and 2 mm erosions to enhance the transition from bright enhanced blood to dark PE.

The three features based on CT value, eigenvalues and bothat are thresholded to create binary masks. To find the optimal threshold for each feature, we used 38 positive data sets with 202 PE that have been confirmed by an experienced chest-radiologist, where discontinuous thrombi were counted as separate lesions. The annotated PE were manually segmented to simplify the matching of detections and annotations. The data sets were acquired with a either a Siemens Sensation 4 or a Philips Mx8000 four-slice CT scanner. The resulting free-response receiver-operator characteristic (FROC) curves are shown in Figure 3a-c. In order to find a good balance between false negatives and false positives, a cost function was defined. We used a function based on the L_1-norm, which is most common. Lines of equal cost are represented in the figure as dashed lines. The lowest cost can be found in the upper-left corner. We have chosen the slope of the dashed lines so that 15 false detections per data set (*FD/ds*) are as costly as 10% missed PE (false negatives). Threshold values for the detectors based on CT value, eigenvalues and bothat were selected and fixed for all data at 0 HU, 1 and 150 respectively to minimize the cost. Variation of the cost function would have led to another optimum (Fig. 3).

The four different detectors respond to different properties of an embolus. The sensitivity is improved by choosing the union of the separate regions. The resulting regions are analyzed using connected-component analysis to obtain the candidates; each group of connected voxels in the combined mask is considered one candidate. Candidates smaller than 8.6 mm^3 are removed, not because they are irrelevant, but to reduce the number of FD. The results of several combinations of detectors can be found in Figure 3d. All four candidates result in a sensitivity of 89% at 41.4 *FD/ds*, which is mainly obtained by the bothat and vessel tracking (3+4, 88% at 39.4 *FD/ds*). So, the two most relevant detectors for our purposes are based on contrast changes (bothat) and unconnected vessels (tracking).



(a) 1: CT value (b) 2: Eigenvalues

(c) 3: Bothat (d) Combined

Fig. 3. (a-c) FROC for the thresholds of three intensity-based detectors and (d) a combination of the four detectors based on 1: CT value, 2: eigenvalues, 3: bothat and 4: vessel tracking. Thresholds of 0 HU, 1 and 150 were chosen for intensity, eigenvalues and bothat respectively to minimize the cost. The dashed lines represent lines of equal costs. The figure shows that the most important detectors are bothat and vessel tracking (3+4).

III. Features for PE Classification

The previous step reduces the search area and creates PE candidates. However, the number of false detections is still too large to serve as output of the CAD system. These false detections are mainly caused by parenchymal diseases, the partial-volume effect on the vessel boundary, sub-optimal contrast enhancement, lymphoid tissue, flow voids in veins, noise and motion artifacts. To make a distinction between real PE and look-alikes, features are computed in candidate objects that allow classification of the candidates.

Only a few attempts have been made to find discriminating features. Masutani e.a. [11] proposed to use the intensity, local contrast, length, volume, curvilinearity of a PE and the vascular size for boundary removal. Zhou e.a. [17], [23] used the features based on intensity, edge strength, length, volume

and the shape of a candidate in relation to the local vessel (roundness and compactness). Pichon e.a. [12] used the size and intensity in a special way. The intensities inside a vessel were projected on the vessel surface by computing the first quartile of intensities on a ray between the surface and the medial axis of a vessel.

Previous work mainly focussed on the intensity and presumed shape of emboli, although other objects can have the same intensity as emboli and emboli can have a wide variety of shapes. Furthermore, the shape of the embolus can be difficult to identify in a CT image, especially if it seems to be connected to the lymphoid tissue (Fig. 1b).

The regular shape of the pulmonary vascular tree (which consist of bifurcations and tubular branches) has not yet been fully exploited. Therefore, we propose to use the shape of a vessel, the shape of lumen (i.e., contrast-enhanced blood) inside these vessels, and the location of a candidate as new features for the classification of PE in CT images.

For each candidate's connected component, we compute the following statistics of a feature: mean (μ), standard deviation (σ), 5th percentile (*min*) and 95th percentile (*max*). For the features that are not measured on the whole object, but for example only on the edge of a candidate, the relative size (*size*) is also computed as a coverage percentage.

A. Intensity

The three intensity-based features that were used for detection are also used for classification. The CT value allows the separation of PE, dark lung parenchyma and bright lumen. The eigenvalues respond to locally dark areas. The bothat transform can in classification be used to remove noise, small flow voids and other areas with a shallow intensity valley.

B. Shape

The first feature based on shape is isophote curvature, which expresses the local shape of a surface through points of equal intensity, and it can be used to measure the shape of bright lumen on its boundary. We propose this shape feature (σ = 0.9 mm) to distinguish the transition 'lumen-tissue' and 'lumen-PE', because PE and tissue cannot be separated based on intensity (Fig. 1b). The lumen surface of healthy pulmonary vessels consists of two shapes; it is ridge-shaped on a tubular branch and saddle-shaped on a bifurcation. The surface shape around an embolus is like a cup or a valley. Unfortunately, the shape measurement of the transitions 'lumen-PE' and 'lumen-tissue' is easily spoiled due to the much larger intensity transition between tissue and parenchyma. Therefore, we want to concentrate only on the relevant intensities. Erf-clipping [24] allows us to do this, while preserving as much edge information as possible (Fig. 4). Isophote curvature allows the detection of the concave lumen surface at embolic locations. However, it will also respond to motion artifacts and at locations where arteries and veins touch each other.

The second feature that expresses the shape of the lumen is its circularity (or eccentricity). The cross-section of healthy pulmonary vessels is circular, but an embolus inside the vessel causes it to become non-circular. Common ways to measure circularity are based on the relation between area and perimeter or the ratio between the eigenvalues of the Hessian ($\frac{|\lambda_2|}{|\lambda_3|}$, where $\lambda_1 \geq \lambda_2 \geq \lambda_3$) [25]. We propose a circularity based on the eigenvalues that is calculated on a distance transform of the lumen segmentation. The distance transform allows the shape analysis of large vessels.

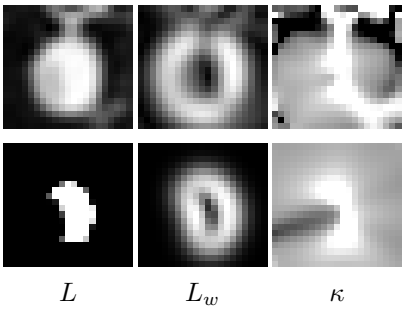


Fig. 4. Intensity (L), gradient magnitude (L_w) and isophote curvature (κ) of a CT image without erf-clipping (top) and of the same image with erf-clipping (bottom). Note the shape of L_w and the negative (dark) values of κ in the erf-clipped images. Erf-clipping allows analysis of the shape of the lumen.

The third feature that uses shape, does not measure the shape of lumen in vessels, but the shape of the vessel. It measures the tubularity of peripheral (segmental and smaller) vessels. Common ways to measure tubularity are vesselness [25] or stringness [26]. We propose to use stringness to measure tubularity near PE candidates, because the ordering of eigenvalues by value (instead of by magnitude) improves the orientation estimation near stenoses and occlusions (like emboli). The stringness is important for PE classification because many emboli can be found in peripheral vessels. Furthermore, most false detections (e.g., caused by tissue or parenchymal diseases) are less tubular than the vessels.

C. Location

The first feature that uses the relative position of a candidate to another structure is based on the distance-to-parenchyma, which is computed with a two-pass distance transform [27] to the region with an intensity below -250 HU. In CTA images, pulmonary vessels are bright and they are surrounded by dark parenchyma. Due to the blurring of the the partial-volume effect (and the PSF) the boundary of a vessel consists of intermediate grey values, which include intensities that are equal to that of emboli. Dark spots inside vessels (and far from the parenchyma) are not caused by the blur effect. Therefore, the distance-to-parenchyma can be used to discriminate emboli from the false detections on the vessel boundary.

The second feature that uses the location of a candidate is based on the distance to the mediastinum in order to indicate whether an embolus is detected in the periphery or near the mediastinum. Therefore, it allows the application of other features in the region for which they are optimized (e.g., stringness for peripheral vessels and isophote curvature in more proximal vessels).

The third feature based on the location of a candidate is the connectivity of a vessel to major vessels, which is generated by the vessel-tracking algorithm (Sec. II-A). Detected vessel parts

that cannot be connected to the mediastinum (unconnected vessels) are marked as suspicious areas.

Not only unconnected vessels are suspected of a complete embolization, but also the tracked paths between an unconnected vessel and the main tree may be important for embolus detection, because it can be caused by PE.

D. Size

A candidate is a group of connected voxels. Its size can be used as a feature to reject false detections. Due to noise and small flow voids dark regions can be detected anywhere. Areas with a shallow intensity valley can in classification be eliminated with the bothat feature, but some noise-related valleys are deep enough to be misclassified as embolus. However, if the size of a candidate is only a few voxels, then the detection of this object might be caused by noise. Therefore, size is also used as a feature to allow removal of small false candidates.

IV. Classification of PE Candidates

The detection step does not only detect true candidates (PE) but also false candidates (look alikes). Classification aims at the separation of these two so that only the true detections will be marked as positives. An embolus is a true positive (TP) when it contains (at least partially, at least one) positive detection, otherwise it is a false negative (FN). Since no classifier performs best in general, we experimentally tested several classifiers on several features. For training and testing, we used 38 positive data sets (202 PE). The data sets were selected to demonstrate a variety of thrombus load, considerable breathing artifacts, sub-optimal contrast and parenchymal diseases.

This section covers feature selection (Sec. IV-A), classifier selection (Sec. IV-B) and performance estimation of the CAD system (Sec. IV-C).

A. Feature Selection

Feature selection is used to find the best set of features corresponding to a classifier. For simplicity, this subsection discusses feature selection only for one classifier: the tree classifier of PRTools [28] (with splitting based on the Fisher criterion and pruning based on the chi-squared test). For the other classifiers we used the same feature selection approach. Those results are presented in Subsection IV-B.

To find the best set of features the technique of *l*-forward and *r*-backward feature selection was used. We tried several initializations, to avoid getting stuck in a local optimum. The forward selection (*l*=2,*r*=1) was not only initialized with an empty set of features, but also with other features and feature pairs. And the backward selection (*l*=1,*r*=2) was not only initialized with all features, but also with the result of forward selection or another chosen group.

We used repeated runs (400) of five-fold cross-validation to estimate the performance, because this is often considered a good trade-off between the bias and variance for the problem of feature and classifier selection [29]–[31].

A criterion for performance estimation is the weighted L_n-norm of the classification error:

$$E = \min(100\,((\alpha\,\tfrac{FP}{ds})^n + (1 - \tfrac{TP}{TP+FN})^n)^{1/n}) \qquad (2)$$

TABLE I
THE L_1-ERROR OF THE BEST SINGLE FEATURES AND FEATURE PAIRS.

Features	L_1-error
Size	69.7
Stringness$_\mu$	72.5
Iso.curv.$_{size}$	82.0
Dist.-to-parench.$_\sigma$	90.4
Stringness$_\mu$ + Dist.-to-parench.$_\sigma$	**41.3**
Peripheral-vessel + Dist.-to-parench.$_\sigma$	50.0
Stringness$_\mu$ + Dist.-to-parench.$_\mu$	54.2
Stringness$_\mu$ + Size	57.5
Eigenvalue$_\mu$ + Dist.-to-parench.$_\sigma$	59.5
Stringness$_\mu$ + Iso.curv.$_{size}$	59.6

with weight $\alpha = 1/30$, which is chosen so that 5 false positives per data set (FP/ds) are as costly as 20% false negatives. The minimum operation ($\min$) refers to the selection of the optimal point on the FROC curve. For (F)ROC analysis, the Manhattan distance (*n*=1) is most commonly used in literature.

Feature selection showed that two features are needed to obtain an L_1-norm smaller than 45: stringness and distance-to-parenchyma. As example, an exploration of the cost of the best single and pairs of features for the tree classifier is shown in Table I. Although others (e.g., size or isophote curvature) also contain valuable information for classification, the best result is obtained with stringness and distance-to-parenchyma. The selection of these two is probably caused by the large amount of PE inside the small tubular vessels and the large amount of false detections on the vessel wall (caused by blur).

B. Classifier Selection

In this subsection, the results of classifier selection and its optimal features are presented. To find a good classifier, feature selection was not only applied to the tree classifier (as in Sec. IV-A), but also to the others, because each classifier may need other features to obtain the optimal result.

We experimentally minimized the L_1-norm on the training data with 400 iterations of 5-fold cross validation for several commonly used classifiers from PRTools [28]. For each classifier we performed feature selection as described in Subsection IV-A. Table II shows the optimal features for each classifier and its corresponding costs. The results show that different classifiers use different features to reach the optimum and that the tree classifier works better than the other used classifiers. Furthermore, they show that the tree classifier uses stringness and distance-to-background to reach the optimum and that other features (such as bothat, size and path-to-unconnected vessels) also contain valuable information.

In our design, the features are optimized for specific properties of true or false detections and their environment (e.g., stringness for small vessels, distance-to-parenchyma for the vessel boundary and isophote curvature for the shape of lumen). Furthermore, the shape of the feature distributions shows a high skew and kurtosis. This type of design and the shape of the distributions indicate that symbolic learning algorithms (such as decision trees) are favored [32].

Bagging improves the classification results for the decision tree. To avoid overtraining, bagging is used in a double loop. In the inner loop, the aggregated tree classifier is trained

TABLE II
COMPARISON OF CLASSIFIERS FROM PRTOOLS [28]. THE DECISION TREE (USING STRINGNESS AND DISTANCE-TO-PARENCHYMA) PERFORMS BEST.

Classifier	Optimal features	L_1
Lin. Bayes norm.	Stringness_μ + $\text{Dist.-to-parench.}_\sigma$	71.8
Lin. PCA	Stringness_μ + $\text{Dist.-to-parench.}_\sigma$	72.4
Lin. log.	BotHat_μ + Size + Stringness_μ + ... Path-to-unconn.	58.2
Lin. discr.	BotHat_μ + Stringness_μ + ... Size + Path-to-unconn.	59.6
Quadr. Bayes norm.	Stringness_μ + $\text{Dist.-to-parench.}_\sigma$	74.3
Quadr. discr.	Size + Unconn.Vessel	72.9
Subspace	EigVal_μ + Iso.Curv_{size} + ... $\text{Dist-to-parench}_\sigma$	65.1
Nearest mean	BotHat_μ + Stringness_μ + Size + Path-to-unconn.	57.9
EM clust. (k=2)	Stringness_μ + Size + ... Path-to-unconn.	60.2
k-NN (k=30)	Stringness_μ + $\text{Dist.-to-parench.}_\sigma$	46.1
Decision tree	**Stringness_μ + $\text{Dist.-to-parench.}_\sigma$**	**41.6**

with bootstrapping and in the outer loop, the performance is estimated with cross validation. As the number of averaged trees (votes) increases, the L_1-error decreases. The bagged tree classifier (32 votes) obtains the same L_1-error with a smaller training set than an unbagged tree classifier.

C. Performance estimation

In this subsection, the performance of the whole system is estimated on the training data: the running time, FROC-curve, analysis of complete occlusions, and the main causes for false positives and false negatives are discussed.

The running time of the system is approximately 30 seconds for one patient on a machine with a Pentium Xeon (3.2 GHz). This is acceptable because the system is expected to assist the radiologist after a manual inspection of the CT data.

The FROC curve is commonly used to present the performance of a classifier when there is an undefined amount of true negatives. Figure 5 shows the FROC curve for classification based on the decision boundary of the bagged tree classifier with the features stringness and distance-to-parenchyma. We used *threshold averaging* to obtain the 68% confidence intervals [33]. A joint confidence region (ellipse) is computed for each point on the curve under the assumption of a bivariate normal distribution based on 400 iterations of 5-fold cross validation on the 38 training data sets. The average performance of the system after classification is estimated to be 70% sensitivity at 5.0 FP/ds, or 75% at 6.5 FP/ds (Fig. 5).

16% of the PE are complete obstructions (present in 30% of the datasets). Our system is able to detect and classify 80% as positive, of which 37% were added by vessel tracking.

The false positives and false negatives were analyzed to gain understanding of the most important causes for misclassification. Each of the misclassifications was assigned at least one cause. The three major causes for FP are: flow voids – or sub-optimal contrast enhancement – in the veins (35%), motion artifacts (29%) and noise in large arteries (13%). Note that the most important causes for false positives of our CAD system (veins and movement artifacts) were not included in the model that was used for the current implementation of

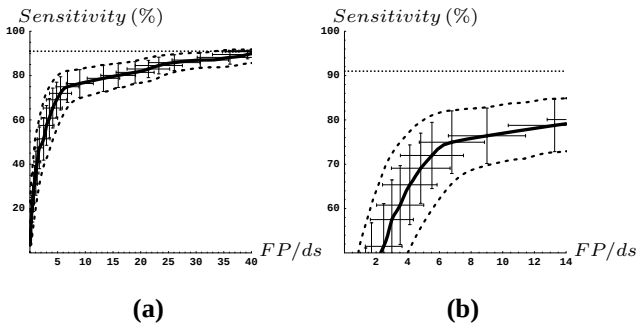


Fig. 5. (a) FROC curve (solid) of the CAD system on the training data with error bars and confidence bands (dashed) based on 68% confidence intervals and (b) detail of this ROC curve.

candidate detection and feature computation. The major causes for FN are: location on the boundary of vessels (65%), motion artifacts (40%) and subtle or tiny PE (40%). The location on the boundary of vessels (related to the feature distance-to-boundary) was often (50%) caused by motion artifacts. In our data, 24% of the PE were tiny or subtle PE and 50% of them were classified as positive.

V. EVALUATION

Our system for the automatic detection of pulmonary embolism in CT images consists of vessel segmentation, candidate detection, feature computation and classification. While tuning the detection system and optimizing the classifier on 38 data sets, the training data may be 'worn out' and the estimated performance can be too optimistic [34]. Therefore, the system was evaluated on 19 positive data sets that were not used in any of the steps described until now. The evaluation results of detection and classification are discussed separately.

A. Detection Results

Initially, 92 PE were annotated by an experienced chest-radiologist in the 19 evaluation sets (Table III). The most proximal location of the thrombus defined its anatomic classification (main, lobar, segmental or subsegmental). In the detection step, 15% (14 PE) were missed by the system at 63 FP/ds. However, the CAD system was also able to find 26% (24 PE) additional PE, which have been confirmed by a radiologist in a second inspection.

The detection step seems to work slightly better for lobar and segmental PE (both 88% sensitivity) than for subsegmental (81%). The number of annotations in the main pulmonary vessels is too low to draw a conclusion from the results. However, based on the design, we know that emboli in the mediastinum are likely to be missed.

B. Classification Results

Figure 5 showed the FROC curve of our system on training data, and Figure 6 shows the same for the evaluation data. The FROC curve on evaluation data shows that the sensitivity of our CAD system is 58%, 63% and 73% at 4.0, 4.9 and 15 FP/ds respectively. These points on the curve are within

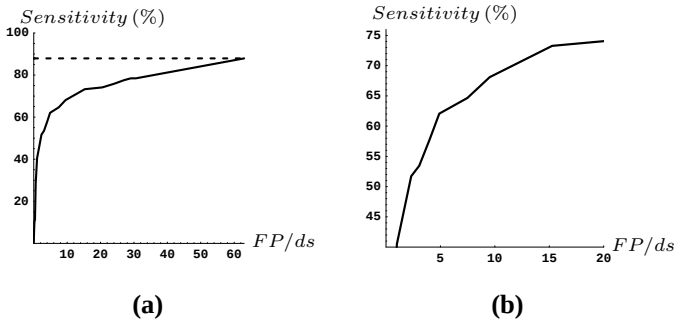
TABLE III
PE ANNOTATED BY THE RADIOLOGIST AND DETECTED BY THE CAD SYSTEM IN THE 19 EVALUATION DATA SETS.

Vessel	Annotated	Detected
Subsegmental	32	26 (81%)
Segmental	34	30 (88%)
Lobar	24	21 (88%)
Main	2	1 (50%)
Initial annotated PE	92	78 (85%)
Additional PE found with CAD	24 (+26%)	24
Total annotated PE	116	102 (88%)
False Positives	–	1197 (63 FP/ds)

a 1.0σ range of our estimation, so the evaluation does not differ significantly from the training results. At 4.9 FP/ds, the system was able to classify correctly: 50% of lobar PE, 65% of segmental and 55% of subsegmental PE (excl. additional).

In the introduction (Sec. I), we already discussed the work and results of others. This discussion is summarized in Table IV for comparison with our results. Masutani e.a. was one of the first who described a computerized method for detecting PE in CTA images. However, it is unclear how well this system generalizes, because it was evaluated on a low number of emboli. The same remark can be made about the system of Pichon e.a. The table also shows that the system of Das e.a. was evaluated on a large number of patients and it performs well in the region where it is important to assist the radiologist. On the other hand, it is unclear how the system will perform on all emboli, because only the results for segmental and subsegmental emboli are reported. Digumarthy e.a. and Maizlin e.a. evaluated the same system. The results of Digumarthy on data without sub-optimal opacification, motion artifacts and diseases are good. However, the evaluation by Maizlin on a small number of more representative data sets showed that the results of the system (both sensitivity and FP/ds) are worse than those of our system.

Only Zhou *et al.*, Buhmann *et al.* and Kiraly *et al.* (apart from us) reported the system performance on a large database of realistic test data that contained artifacts and diseases without the exclusion of PE. We are able to obtain a higher sensitivity at a lower number of FP/ds than the detection systems of Zhou and Buhmann. At only 4.9 FP/ds, we are able to obtain a higher sensitivity than the visualization method of Kiraly and a higher additional value (22% added PE).



(a) **(b)**

Fig. 6. *(a)* FROC curve of the CAD system on the evaluation data and *(b)* detail of this FROC curve.

TABLE IV
COMPARISON OF SYSTEMS FOR PE DETECTION. THE TABLE SHOWS THE NUMBER OF PE, THE NUMBER OF POSITIVE DATA SETS (ds) USED FOR EVALUATION, SENSITIVITY (Sens, %), FP/ds, AND A REMARK.

Reference	PE	ds	Sens	FP/ds	Remark
Masutani [11]	21	11	100	7.7	Low #PE.
Pichon [12]	22	3	86	6.3	Low #PE.
Das [13]	168 seg 120 sub	33	88 78	4	Only Peripheral PE.
Digumarthy [14]	270	39	92	2.8	No motion; No diseases
Maizlin [15]	45	8	58	6.4	Low sens. %; Low #ds
Kiraly [18]	69	8	50	–	13% add. PE.
Zhou [17]	225	14	52	11.4	Low sens. %
Buhmann [16]	352	40	47	3.9	Low sens. %
Proposed system	116	19	58 63 73	4.0 4.9 15	Artif. & disease; 22% add. PE. (see Fig. 6)

VI. DISCUSSION

With our 63% sensitivity CAD system, we are able to add 22% to the manually detected PE of the radiologist. However, we do not know the relevance of the additional emboli for two reasons. Most of the additional PE were found in the segmental and subsegmental (peripheral) pulmonary arteries. However, the clinical relevance of peripheral or small PE is a matter of controversy [9], [35], [36]. Although this relevance is still unclear, we believe that a higher sensitivity will lead to a more accurate estimation of the severity of obstruction and a better patient selection. The second reason for unclarity about the relevance of additional emboli is related to positive training and evaluation data. Our results were obtained on a group of positive data sets, which contained (on average) five PE per data set. However, for treatment, it is less relevant to find an extra embolus in a data set that has already been classified as positive, than to find an embolus in a data set that would otherwise have been classified as negative. Unfortunately, the estimated added value of our CAD system cannot easily be extrapolated to data sets that were initially classified as negative, because the sensitivity of the radiologist – which influences the added value – might not be constant. Thus, based on the results, it is difficult to estimate the clinical relevance of our CAD system. Although the system can find additional emboli and the performed evaluation is at the level of state of the art literature, extension of the evaluation with both positive and negatives scans would allow an analysis of the effects of our system on patient outcome.

VII. CONCLUSIONS AND RECOMMENDATIONS

In this paper, we showed that our system for the automatic detection of PE candidates was able to find approximately 90% of the emboli at 41 false detections per data set. Of the four features that were used for candidate detection – which were based on CT values, eigenvalues of the Hessian, bothat transform and vessel tracking – the last two appeared to improve the detection results the most. This shows that contrast changes (bothat) and unconnected vessels (tracking) are important in the detection step.

We also proposed new features to be used for the classification of pulmonary embolism. We did not only focus on characteristics of the embolus, but also on features that describe the blood vessels. The features are based on intensity, location, size, shape of lumen and shape of a vessel.

We showed that the bagged tree classifier – with the features 'distance-to-parenchyma' and 'stringness' – optimizes the performance of the system. Other features also contained valuable information but they were not able to improve the end-result. The selection of these two features may be related to a large amount of PE inside the small tubular vessels and a large amount of false detections on the vessel wall (caused by the partial-volume effect).

Our system performs well in comparison to other CAD systems presented in literature that were evaluated extensively in a comparable manner. The sensitivity of the CAD system is 63% at 4.9 false positives per data set, which allowed the radiologist to improve the number of detected PE by 22%.

The most important causes for false positives are: flow voids in veins, motion artifacts and noise due to sub-optimal contrast. The most important causes for false negatives are: location on the boundary of a vessel, motion artifacts and subtle or tiny PE. Only a small number of misclassifications was caused by tissue near the mediastinum, parenchymal diseases or complete occlusions. Apparently our system is able to handle these causes successfully.

Future research may include further analysis of the difference between PE in large proximal and small peripheral vessels and it may focus on the segmentation of whole emboli. An artery-vein separation and the recognition of motion artifacts are also an interesting field for future research, which can reduce the number of misclassifications. However, these types may be automatically removed (or become less important) when newer CT scanners or improved scanning protocols are used. Another interesting topic would be to use stringness and distance-to-background in the segmentation of the search area, which would have excluded many false detections. This might have allowed the use of other features in classification to improve the end-result. Furthermore, the evaluation could be extended to a large number of positive and negative scans to see whether the CAD system will change patient outcome.

ACKNOWLEDGMENTS

The authors thank Dr. C.M. Schaefer-Prokop and Dr. O.M. van Delden of the AMC in Amsterdam for the annotated CTA data, and J. Peters of Philips Healthcare for the discussions.

REFERENCES

[1] A. Pforte, "Epidemiology, diagnosis and therapy of pulmonary embolism," *European J. Medical Research*, vol. 9, pp. 171–179, Apr. 2004.

[2] J. Juni and A. Abass, "Lung scanning in the diagnosis of pulmonary embolism," in *Proc. Nucl. Med.*, vol. 21, 1991, pp. 282–296.

[3] J. Ravenel, H.P. McAdams, P.C. Goodman, "Computed tomography pulmonary angiography: Diagnostic pitfalls and artifacts," *Applied Radiology*, vol. 30, no. 10, pp. 35–42, 2001.

[4] M. Domingo, L. Martí-Bonmatí, R. Dosdá and Y. Pallardó, "Interobserver agreement in the diagnosis of pulmonary embolism with helical CT," *European Journal of Radiology*, vol. 34, pp. 136–140, 2000.

[5] I. Hartmann and M. Prokop, "Spiral CT in the diagnosis of acute pulmonary embolism," *Medica Mundi*, vol. 46, no. 3, pp. 2–11, 2002.

[6] M. Remy-Jardin and J. Remy, "Spiral CT angiography of the pulmonary circulation," *Radiology (RSNA)*, vol. 212, pp. 615–636, Sep. 1999.

[7] N. Schibany, D. Fleischmann, C. Thallinger and others, "Equipment availability and diagnostic strategies for suspected pulmonary embolism in Austria," *European Radiology*, vol. 11, no. 11, pp. 2287–2294, 2001.

[8] C. Fink, S.O. Schoenberg and M.F. Reiser, "MRI expands options in lung assessment," *Diagnostic Imaging Europe*, pp. 14–19,47, Nov. 2005.

[9] U. Schoepf, M.A. Kessler, C.T. Rieger e.a., "Multislice CT imaging of pulmonary embolism," *Eur. Rad.*, vol. 11, no. 11, pp. 2278–2286, 2001.

[10] P. Stein, S.E. Fowler, L.R. Goodman e.a., "Multidetector computed tomography for acute pulmonary embolism," *New Engl. J. Med.*, vol. 354, no. 22, pp. 2317–2327, 2006.

[11] Y. Masutani, H. MacMahon, K. Doi, "Computerized detection of pulmonary embolism in spiral CT angiography based on volumetric image analysis," *IEEE T. Med. Im.*, vol. 21, no. 12, pp. 1517–1523, 2002.

[12] E. Pichon, C.L. Novak, A.P. Kiraly and D.P. Naidich, "A novel method for pulmonary emboli visualization from high-resolution CT images," in *Proc. SPIE Medical Imaging*, vol. 5367, 2004, pp. 161–170.

[13] M. Das, A. Schneider, U. Schoepf, e.a., "Computer-aided diagnosis of peripheral pulmonary emboli," in *Proc. RSNA*, no. C02-232, 2003.

[14] S. Digumarthy, C. Kagay, A. Legasto, e.a., "Computer-aided detection (CAD) of acute pulmonary emboli: Evaluation in patients without significant pulmonary disease," in *Proc. RSNA*, no. SSC04-08, 2006.

[15] Z. Maizlin, P. Vos, M. Gody and P. Cooperberg, "Computer-aided detection of pulmonary embolism on CT angiography: Initial experience," in *Proc. Ann. Meeting RSNA*, no. SSC04-04, 2006.

[16] S. Buhmann, P. Herzog, J. Stoeckel, M. Salganicoff, M. Wolf, M.F. Reiser and C.R. Becker, "Clinical evaluation of a CAD prototype for the detection of pulmonary embolism," in *ECR*, 2006.

[17] C. Zhou, H.-P. Chan, S. Patel, e.a., "Preliminary investigation of computer-aided detection of pulmonary embolism in 3D CT pulmonary angiography," *Ac. Rad.*, vol. 12, no. 6, pp. 782–792, 2005.

[18] A. Kiraly, C.L. Novak, D.P. Naidich, e.a., "A comparison of 2D and 3D evaluation methods for pulmonary embolism detection in CT images," in *Proc. SPIE Med. Im.*, vol. 6146, 2006.

[19] J. Peijl, "Automatic detection and segmentation of the lungs in CT datasets," Master's thesis, TU Eindhoven, 2003, http://bmia.bmt.tue.nl.

[20] S. Ukil and J.M. Reinhardt, "Smoothing lung segmentation surfaces in 3D X-ray CT images using anatomic guidance," in *Proc. SPIE Medical Imaging*, vol. 5370, 2004, pp. 1066–1075.

[21] T. Bülow, R. Wiemker, T. Blaffert, C. Lorenz and S. Renisch, "Automatic extraction of the pulmonary artery tree from multi-slice CT data," in *Proc. SPIE Medical Imaging*, vol. 5746, 2005, pp. 730–740.

[22] P. Soille, *Morphological Image Analysis*. Springer-Verlag, 1999.

[23] C. Zhou, L.M. Hadjiiski, B. Sahiner, e.a., "Computerized detection of pulmonary embolism in 3D CT images: vessel tracking and segmentation techniques," in *Proc. SPIE Med. Im.*, vol. 5032, 2003, pp. 1613–1620.

[24] L. Vliet and P.W. Verbeek, "Better geometric measurements based on photometric information," in *IEEE Proc. Instrumentation and Measurement Technology Conf.*, vol. 3, Japan, May 1994, pp. 1357–1360.

[25] A. Frangi, W.J. Niessen, K.L. Vincken, M.A. Viergever, "Multiscale vessel enhancement filtering," in *MICCAI*, ser. LNCS, vol. 2489, 1998.

[26] P.-E. Danielsson, Q. Lin and Q.-Z. Ye, "Efficient detection of second-degree variations in 2D and 3D images," *J. Vis. Comm. Im. Representation*, vol. 12, no. 3, pp. 255–305, 2001.

[27] A. Rosenfeld and J.L. Pfaltz, "Sequential operations in digital picture processing," *J. ACM*, vol. 13, no. 4, pp. 471–494, Oct. 1966.

[28] R. Duin, P. Juszczak, D. de Ridder e.a., "PRTools, a Matlab toolbox for pattern recognition," TU Delft, Tech. Rep., 2004, http://www.prtools.org.

[29] T. Bailey and C. Elkan, "Estimating the accuracy of learned concepts," in *Proc. Int. Joint Conf. Artificial Intelligence*, 1993, pp. 895–900.

[30] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. AI*, 1995, pp. 1137–1145.

[31] B. Efron and R. Tibshirani, "Improvements on cross-validation: The .632+ bootstrap method," *J. Am. Stat. Assoc.*, vol. 92, no. 438, 1997.

[32] R. King, C. Feng and A. Sutherland, "STATLOG: Comparison of classification algorithms on large real-world problems," *Applied Artificial Intelligence*, vol. 9, no. 3, pp. 289–334, 1995.

[33] S. Macskassy and F. Provost, "Confidence bands for ROC curves: Methods and an empirical study," in *ROC Analysis in AI*, 2004.

[34] Q. Li and K. Doi, "Some practical issues for assessment of computer-aided diagnostic schemes," in *SPIE Medical Imaging*, vol. 6514, 2007.

[35] L. Goodman, "Small pulmonary emboli: What do we know?" *Radiology (RSNA)*, vol. 234, no. 3, pp. 654–658, March 2005.

[36] J. Wildberger, A.H. Mahnken, M. Das, e.a., "CT imaging in acute pulmonary embolism," *Eur. Rad.*, vol. 15, no. 5, pp. 919–929, 2005.