Detection of Pulmonary Nodules in Low-dose Computed Tomography Using Localized Active Contours and Shape Features

Abstract

Background: Pulmonary nodules are symptoms of lung cancer. The shape and size of these nodules are used to diagnose lung cancer in computed tomography (CT) images. In the early stages, nodules are very small, and radiologist has to refer to many CT images to diagnose the disease, causing operator mistakes. Image processing algorithms are used as an aid to detect and localize nodules. Methods: In this paper, a novel lung nodules detection scheme is proposed. First, in the preprocessing stage, our algorithm segments two lung lobes to increase processing speed and accuracy. Second, template-matching is applied to detect the suspicious nodule candidates, including both nodules and some blood vessels. Third, the suspicious nodule candidates are segmented by localized active contours. Finally, the false-positive errors produced by vessels are reduced using some two-/three-dimensional geometrical features in three steps. In these steps, the size, long and short diameters and sphericity are used to decrease the false-positive rate. Results: In the first step, some vessels that are parallel to CT cross-plane are identified. In the second step, oblique vessels are detected using shift of center of gravity in two successive slices. In step three, vessels vertical to CT cross-plane are identified. Using these steps, vessels are separated from nodules. Early Lung Cancer Action Project is used as a popular dataset in this work. Conclusions: Our algorithm achieved a sensitivity of 90.1% and a specificity of 92.8%, quite acceptable in comparison to other related works.

Keywords: Computer-aided detection, computed tomography images, feature extraction, localized active contours, pulmonary nodules, template matching

Introduction

Lung cancer in the early stages has no symptoms in serial computer tomography (CT) images, whereas its early detection can increase survival rate. The symptoms of lung cancer only appear in more serious stages due to characteristics and position of the nodules in the lungs. Lung nodules are round- or egg-shaped lesions with regular or irregular boundaries. Nodules in early stages are very small, thus high-quality images need to be used, and these images need to be processed by a radiologist visually. In such a situation, computer-aided detection (CAD) systems process the CT images to assist the radiologist in detecting pulmonary nodules and increasing sensitivity and specificity. Suspicious areas are mainly due to blood vessels which resemble nodules.

In the remainder of this section, two categories of papers are investigated. The first category discusses nodule detection,[2-16] and the second group considers nodule segmentation.[17-23] Murphy et al.[2] described a scheme for the automatic detection of nodules in thoracic CT scans that enhances false-positive reduction using two successive k-nearest neighbor classifiers. Liu et al.[3] proposed a computer-aided diagnosis system for lung nodule detection using the hidden conditional random field. This method achieved sensitivity of 89.3%. Keserci and Yoshida[4] proposed a new computer-aided diagnosis scheme for automated detection of lung nodules in digital chest radiographs based on combination of morphological features and wavelet snake model. Choi and Choi[5] presented an algorithm based on two-dimensional (2D) and three-dimensional (3D) features to reduce false-positive errors. This system first enhanced the CT images and then detected nodule candidates. Afterward, the features are extracted from every region of interest (ROI). The 3D geometric feature...
set consisted volume elongation factor, compactness, and approximated radius. In addition, the 2D features were extracted from middle slice of nodule candidates. These features were mean, variance, skewness, kurtosis, area, radius, and eight biggest eigenvalues. Finally, these features were applied as input to a genetic programming module. The method applied to 32 scans consisting of 153 nodules and 7528 slices. Half of dataset (16 scans) was used for training, and the remaining was used to test the classifier. This method achieved sensitivity of 92% with an average of 6.5 false-positives per scan. Suzuki et al. investigated a pattern-recognition technique based on an artificial neural network for the reduction of false positives of detected lung nodules in low-dose CT.

Gao et al. proposed a novel template-matching using Hessian matrix in a series of chest X-ray images. First, the lung lobes were extracted in the original images. Then, template-matching method was used to detect the suspicious nodule candidates. Although it was fast and nearly all nodules were detected, it had a high false-positive rate. Hence, to classify nodule candidates into two groups, nodules and vessel, a classifier based on eigenvalues of Hessian matrix was proposed. The algorithm depended on a few constants that had to be specified experimentally. This method was applied to both a high-resolution CT and to a normal CT. Yogananda et al. proposed an algorithm to detect the pulmonary nodules. After preprocessing step, morphological operations were used to detect the nodule candidates. Then, to discriminate between the nodules and the vessels, the successive frames of the CT were examined. If the candidate was a vessel, the displacement of center of gravity of the candidate in successive frames was large. However, if the candidate was a nodule, the displacement would be small. This algorithm failed to separate the nodules from vertical vessels. Namin et al. developed an automatic method for lung nodule detection. After preprocessing step, suspicious nodule candidates were detected on the basis of intensity and volumetric shape index (SI). This method achieved sensitivity of 88% with an average of 10.3 false positives per scan. Lin et al. developed a neural network system based on a two-level convolution neural network architecture to reduce false positives. Shi et al. employed neural network ensemble for false positives reduction in detecting lung nodules in chest radiographs. The performance of this algorithm was evaluated by use of FROC. Dolejsi et al. designed a CAD system both to detect small pulmonary nodules efficiently and to reduce false positives based on Asymmetric Adaboost ensemble. Guo et al. proposed a scheme on the basis of an adaptive classification system. Eight features were extracted from nodule candidates, and support vector machine (SVM) was used as a nonlinear classifier. Assefa et al. proposed an algorithm using template-matching and multiresolution analysis to improve false-positive rate. In this algorithm, intensity and multiresolution features were used. This method achieved detection rate of 81.212% and false-positive rate of 35.15%. Farag et al. determined the models for lung nodule detection by template-matching method. This algorithm emphasized on the features using texture and shape properties of nodules. de Carvalho Filho et al. proposed an algorithm based on taxonomic diversity and taxonomic distinctness indexes from ecology to describe the texture of nodule and nonnodule candidates. Then, SVM was used as a classifier. This algorithm achieved a mean accuracy of 98.11%.

By considering the abovementioned papers, there are generally two approaches in nodule detection: detection using only a single 2D frame and detection using a series of 2D frames. As it is heuristically clear, experiments show that detection results are much better, when multiple frames are used for diagnosis. Furthermore, the experimental results have shown that accurate segmentation of regions in CT images has a significant effect in the diagnosis of pulmonary nodules. Therefore, many algorithms have been developed for this purpose. In the following, a brief explanation of algorithms used for lung nodule segmentation is presented.

Messay et al. used a regression neural network to determine threshold parameters in lung nodule segmentation. They tested their algorithm on 66 lung nodules and achieved a segmentation accuracy of 80%. Qian and Guirong proposed the expectation maximization (EM) algorithm for lung nodule detection and segmentation. Their model was independent from the size of nodules. Sun et al. developed an automatic segmentation algorithm of nodules in CT images. This method successfully distinguished nodules attached to pleural surface, but it failed to segment the cavity nodules. Zinoveva et al. proposed a soft segmentation based on pixel-level texture features using a decision-tree classifier with classification and regression tree algorithm. This method successfully distinguished the nodules that were attached to vessels, but it failed to segment those attached to the chest wall. To solve this problem, the lung segmentation was performed in the preprocessing step. Okada and Akdemir proposed a segmentation method using SI. This algorithm provided successful results to segment the spherical nodules, but it failed to detect the nonspherical nodules. The model used in this paper successfully segmented ground-glass opacity (GGO) nodules, cavity nodules, and nodules that are attached to vessels or chest wall.

To achieve a high-performance level for lung nodules detection, an efficient method employs three major steps: initial nodule detection, segmentation process, and false-positive reduction. In this study, we employed a template matching (TM) method to increase the detection sensitivity. Then, we utilized a local active contour to
increase segmentation accuracy; this method is able to segment challenging nodule candidates (vessels, different types of nodules including GGO nodules, cavity nodules, spherical and nonspherical nodules, and nodules that are attached to vessels or chest wall) that some previous works fail. Finally, we proposed a three step false-positive reduction algorithm; this approach was used to distinguish nodules from other suspicious regions including oblique blood vessels and vessels that are parallel or vertical to CT cross-plane. The main novelties of this work are automated nodule detection along with discriminating nodules from different type of vessels using shape features, which leads to a high sensitivity and specificity. Figure 1 gives an overall overview of our scheme for nodule detection in CT images. Accordingly, the main novelties of this work are automated nodule detection along with discriminating nodules from different type of vessels using shape features, which leads to a high sensitivity and specificity. Figure 1 gives an overall overview of our scheme for nodule detection in CT images. Accordingly, the remainder of this paper continues as follows: our proposed algorithm is presented in Section 2. In Section 3, we discuss the experimental results. Conclusion is given in Section 4.

Materials and Methods

Preprocessing process

Nodule detection process is divided into the following steps: preprocessing, identification of suspicious nodule candidates, classification, and false-positive reduction. In the preprocessing stage, the lung lobes are segmented in CT images. The approach proposed in [24] is used to segment the lung from its surrounding based on EM algorithm and morphological operations. Furthermore, contrast adjustment is applied to improve the quality of the lung lobe images. A sample is shown in Figure 2.

Detection of suspicious nodule candidates

TM can be used to detect the pulmonary nodules efficiently, but at the same time many vessels are also detected as nodules in this manner.[7] Intensity value of nodules usually has a Gaussian-like distribution as follows:[15]

\[ q(r) = q_{\text{max}} \cdot e^{-r^2/\rho^2}, \quad 0 \leq r \leq R \]

(1)

Where,

\[ \rho = R \left( \ln(q_{\text{max}}) - \ln(q_{\text{min}}) \right)^{0.5} \]

In which \( q_{\text{min}} \) and \( q_{\text{max}} \) are the minimum and maximum of nodule intensity, \( q(r) \) is the intensity of nodule at distance from the centroid of the nodule, and \( R \) is the radius of nodular template.

TM method computes the normalized cross-correlation (NCC) of the image and the template by the following equation:

\[ \gamma = \frac{\sum_{x,y} [f(x,y) - \bar{f}_{xy}][t(x-u,y-v) - \bar{t}]}{\sqrt{\sum_{x,y} [f(x,y) - \bar{f}_{xy}]^2 \sum_{x,y} [t(x-u,y-v) - \bar{t}]^2}} \]

(2)

Where

\[ \bar{f}_{xy} = \frac{1}{N_x N_y} \sum_{x'=0}^{N_x-1} \sum_{y'=0}^{N_y-1} f(x',y') \]

\[ f(x,y) \] and \( \bar{f}_{xy} \) are the input image and its mean value in the region under template located at \( (u,v) \), \( t(x,y) \) is the nodule model and its mean value is \( \bar{t} \). In (2), the correlation coefficient \( \gamma \) varies between 1 and −1. If the correlation coefficient is \( \gamma \) above a positive threshold level, the point \( (u,v) \) is accepted to belong a suspicious nodule candidate.[7] This threshold level according to our database is experimentally set to 0.6.

In our experiments, we used a library of circular and semi-circular templates with 90° rotations with a Gaussian distribution of grayscale intensity. The mean diameter of nodules in our database is 8.5 mm with a standard deviation of 3.6 mm. Accordingly, we chose a diameter of 8.5 mm for our circular and semi-circular templates the same as.[15]

Using this method, many regions are tagged, but the true positive rate is high. Figure 3 shows a sample of this method. The template image is seen as Figure 3a. After removed interference, the pulmonary parenchyma image is gotten by best weight segmentation and seen as Figure 3b. The normalized cross-correlation of the image and the template is seen as Figure 3c and the initial regions of interest are shown as Figure 3d.

It should be noted that in the above procedure, some vessels are also detected as suspicious nodule candidates.
2D processing is frequently not able to remove the interference of vascular, leading to high false-positive rate. To separate the nodules from vessels in the next section, nodule candidates are segmented, and then these segmented nodule candidates are classified using some features.

**Segmentation process**

Accurate segmentation of suspicious nodule candidates, that could be a nodule or a vessel, is very important. Segmentation of nodule candidates is challenging because of the following three effects: noise, nodules that are attached to blood vessels or attached to lung wall, and low contrast of intensity values between nodules and other structures in CT images, due to low-dose CT imaging.

The basic idea in active contour models is to allow a contour to deform so as to minimize a given cost function to provide the desired segmentation results. In general, there are two main categories of active contour models: edge-based and region-based.

Edge-based active contour models mainly utilize the edge and gradient information to drive the contours to identify object boundaries. The result of image segmentation by these models is highly dependent on the initial contour placement and very sensitive to image noise.

Compared with edge-based active contours, region based active contours model the foreground and background regions statistically and optimize a global energy cost function. These models are less sensitive to initialization and image noise. Region-based active contours focus on a localized energy function based on piece-wise constant model of Chan and Vese (C-V model), which can be written as:

\[
F(u,v,\phi) = \mu \int_{\Omega} \nabla H(\phi) \cdot \nabla x + \nu \int_{\Omega} H(\phi) \, dx \\
+ \lambda_1 \int_{\Omega} (H(\phi)(I-u)^2 \, dx + \lambda_2 \int_{\Omega} (1-H(\phi))(I-v)^2 \, dx
\]

Where \( I \) is a given image defined on domain \( \Omega \), \( \nu \), and are \( \nu \) the global mean intensities of the interior and exterior regions. \( \lambda_1, \lambda_2, \nu \) and \( \mu \) are positive constant coefficients and is \( H(\phi) \) Heaviside function.

Optimizing global statistics usually is not ideal for segmenting heterogeneous objects. Therefore, to accurately segment these objects, a new model of active contour is developed which utilizes local information.

Localized active contours are capable of segmenting objects with heterogeneous feature profiles. In this approach, segmentation is not based on global region models. The average intensities in interior and exterior areas of a mask \( B(x,y) \) are computed at each point. To optimize the total energy of the contour, the mask is considered in each point separately, and the point is moved to decrease the energy function. This energy function is defined as following:

\[
E(\phi) = \int_{\Omega} \delta(\phi(x)) \int_{\Omega} B(x,y) F(I(y),\phi(y)) \, dx dy \\
+ \lambda \int_{\Omega} \delta(\phi(x)) \| \nabla \phi(x) \| \, dx
\]

\( F \) is a generic internal energy, and \( \delta(\phi(x)) \) is Dirac function. Using \( B(x,y) \), \( F \) operates only on local image information in neighborhood of \( (x,y) \) and \( \lambda \) is a smoothing parameter. Using calculation of variation results:

\[
\frac{\delta E(\phi)}{\delta \phi} (x) = \delta(\phi(x)) \left[ (B(x,y)) \, \delta(\phi(y)) - \lambda \int_{\Omega} \delta(\phi(x)) \| \nabla \phi(x) \| \, dx \right]
\]

The localized equivalents of and that defined in terms of the \( B(x,y) \) function, \( u \) and \( v \):

\[
u = \frac{\int_{\Omega} B(x,y)H(\phi(x))I(y)dy}{\int_{\Omega} B(x,y)H(\phi(x))dy}
\]

\[
u = \frac{\int_{\Omega} B(x,y)H(\phi(x))I(y)dy}{\int_{\Omega} B(x,y)H(\phi(x))dy}
\]

Figure 2: Lung lobe segmentation with contrast adjustment. (a) Input image, (b) mask of segmented lung region, (c) segmented lung

Figure 3: Suspicious nodule candidate detection. (a) Template, (b) original image, (c) normalized cross-correlation of image and template, (d) result of template matching method
Nadealian, et al.: Detection of pulmonary nodules

Segmented nodule candidate regions from adjacent structures in majority of cases but failed to discriminate between background and some vessels. Figure 6 shows a vessel segmented by localized active contour. As it is shown in this figure, some gaps are generated within vessels. It should be noted that such a failure makes no problem in the final results because vessels are identified from nodules and omitted in later stages.

**Discriminating blood vessels from nodules**

In each CT cross-section, some vessels may resemble nodules, causing false nodule candidates which must be identified and omitted. The following three steps do the job. Nonnodule objects can be discriminated from nodule candidates using information of nodules such as size, volume, sphericity, and compactness.

**Detection of vessels parallel to computed tomography cross-plane**

Geometrical shapes of pulmonary nodules and vessels are spherical and cylindrical, respectively. A slice in a sphere, semi-sphere, vertical, or nearly vertical cylinder is circular or nearly circular. As the cylinder inclines more parallel to CT cross-plane, the difference between the long diameter and the short diameter increases. The long diameter is the distance between two points of nodule boundary with maximum distance, and the short diameter is the longest chord perpendicular to the long diameter.

We use long diameter/short diameter as a feature to classify vessels and nodules. The ratio is large for oblique vessels or vessels vertical to CT cross-plane and is small for nodules. The diagrams in Figure 7 show the distribution of long diameter/short diameter for nodules and vessels. The diagram vertical axis is the number of nodules/vessels with specific long diameter/short diameter of horizontal axis. The threshold value is experimentally set to 2.1. Some of the vessels that are detected or not detected in this step

\[
A(x, y) = \int_{\Omega} I(x,y) \cdot (1 - H\phi(y))dy
\]

Local active contour has three parameters, maximum iterations which is set to 200, local radius \(B(x,y)\) set to 1 and smoothing parameter \(\lambda\) set to 0.2. The parameters are chosen according to nodule specifications experimentally.

Figure 4 shows the results of global and local active contours applied to segmentation of a nodule. It can be seen that local active contour successfully segmented the nodule candidates with concave and low contrast boundaries.

This model was tested on different kinds of nodule candidates including both nodules and vessels. The experiments show the robustness and reliability of the model. As shown in Figure 5, this method successfully

![Figure 4: (a and c) Initialization, (b) final result with global energies, (d) final result with local energies](image)

![Figure 5: Segmentation results. (a and c) are nodule and (e and g) are vessel that are shown by blue arrows, (b, d, f and h) are the contour of the segmentation](image)
are shown in Figure 8. In the next steps, remaining nodule candidates are examined to detect other possible vessels.

**Detection of oblique blood vessels**

If we consider the center of gravity of a nodule or vertical vessel in two successive slices, there would be a small displacement between these two centers [Figure 9], whereas this displacement would be greater for an oblique vessel. The diagrams in Figure 10 show the distributions of these displacements for nodules and vessels. Hence, a number of oblique blood vessels in nodule candidates can be identified using a threshold on these displacements. The diagram vertical axis is the number of nodules/vessels with specific displacements of horizontal axis. The threshold value is experimentally set to 0.75.

**Detection of vessels vertical to computed tomography cross-plane**

In the previous two steps, oblique and parallel vessels were removed from the suspicious nodule candidate list. In the current step, we will only consider the remaining suspicious nodule candidates. A vessel is not a compact object and continues to be connected in consecutive slices of CT images. On the contrary, a nodule as shown in Figure 11 is a compact and sphere object, so its cross-section on a CT slice is nearly a semi-circle.

Experimental results indicated that there is a small difference between the length and short diameter of a nodule, whereas the difference would be higher for vascular areas because vascular areas continues to be connected in successive slices of 2D CT images. Thus, the length and short diameter of remaining suspicious nodule candidates are extracted as features. Furthermore, the length of nodule candidates is calculated as follows:

\[
\text{Length} = n \times \text{slice thickness}
\]

\(n\) is the number of slices that transverses a nodule candidate. In this study, slice thickness of data is mm. To determine, the following algorithm is applied as in Figure 12:

1. Three slices, the current slice and slices before and after it are considered
2. The cross-section area of nodule is already segmented by local active contour
3. Its relative area in previous and next frames are marked, and the brightest pixels in these marked areas are determined and labeled by points \(P_1\) and \(P_2\), respectively
4. \(P_1\) and \(P_2\) are selected as the center of the initial local active contours
5. The radius of the active contours is 2.

The following cases may occur as shown in Figure 12b and c.

1. Each contour in previous/next slice has just one region. \(M_{p1} \mid M_{p2}\) is the center of gravity in this region. If the following two conditions are fulfilled, the candidate area is considered to belong to the current nodule candidate
   a. The average intensity in this region be greater than a threshold
   b. The displacement of its center of gravity with respect to the center of gravity of the current slice, shown with vectors \(d_{p1} \mid d_{p2}\) be smaller than a threshold
2. If the region segmented in the previous/next slice has multiple areas or has just one pixel, as in Figure 12c,
According to the above algorithm, the total number of slices belonging to the nodule \( n \) and so its length is determined.

The diagrams in Figure 13 show the scatter plot of the short-axis diameter and length for nodules and vessels. Using the features extracted from the regions, the length and short-axis diameter of the nodule candidates, SVM is used to classify the nodules and vessels. In experimental results, it is shown that SVM-RBF improves the performance of classification. This classifier is evaluated by 5-fold cross-validation.

**Experimental Results**

In this paper, we use the Early Lung Cancer Action Program public database, which is one of the lung image database consortium groups. This database contains fifty sets of low-dose CT lung scans taken at a single breath-hold with slice thickness 1.25 mm. The locations of nodules were certified by four radiologists. Accordingly, 31% of nodules are solitary, 30% are bronchiole attached, and 39% are lung wall attached, where 39.12% are juxtapleural nodules, 13.95% are vascularized nodules, 31.29% are well-circumscribed nodules, and 15.65% are pleural-tail nodules. The mean diameter of nodules in this database is 8.5 mm with standard deviation of 3.6.\(^{[28]}\) It includes 50 sets of low-dose CT scans with a total of 12645 slices. CT scan images of 20 patients were used to develop the algorithm, and the remaining images were used to test the algorithm. The number of slices for a patient varies from 212 to 304. Slice thickness is 1.25 mm in this database, and pixel spacing is from 0.5078 mm to 0.8164 mm for different patients.

At first, we used TM method to find nodule candidates, and then we segmented the area of nodule candidates using localized active contours. Then most of the blood vessels were removed from the nodule candidate list according to the three steps in section 2.4. The parameters of all steps of the algorithm are chosen experimentally according to nodule specifications in CT images of our database.
Sensitivity and specificity of our algorithm were 90.1% and 92.8%, respectively. Finally, the results were compared with some of the latest similar researches on the subject as shown in Table 1. In this table, the highest sensitivity is 96.15% in \[29\] which is better than 90.1% of our algorithm, but our specificity is 92.8, better than 52.17% in \[29\]. This result shows that, compared with existing algorithms, our method performs at a similar or better level.

The receiver operating characteristic (ROC) curve is shown in Figure 14. ROC analysis was used to evaluate the automated classifier. The area under the ROC curve is a

<table>
<thead>
<tr>
<th>Method</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Database</th>
<th>Number of cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Farag et al.[15]</td>
<td>85.22</td>
<td>86.28</td>
<td>ELCAP</td>
<td>50</td>
</tr>
<tr>
<td>Orozco et al.[29]</td>
<td>96.15</td>
<td>52.17</td>
<td>ELCAP</td>
<td>75</td>
</tr>
<tr>
<td>Assefa et al.[14]</td>
<td>81.212</td>
<td>64.85</td>
<td>ELCAP</td>
<td>50</td>
</tr>
<tr>
<td>Our algorithm</td>
<td>90.1</td>
<td>92.8</td>
<td>ELCAP</td>
<td>30</td>
</tr>
</tbody>
</table>

NBIA – National biomedical imaging archives; ELCAP – Early Lung Cancer Action Project

Figure 11: Feature parameters of a nodule

Figure 12: (a) Active contour of current slice and the two corresponding regions in previous and next slices. (b) Single regions segmented by local active contours in previous and next slices. (c) Multiple regions or one pixel

Figure 13: Scatter plot for short diameter versus length for (a) nodules, (b) vessels
good measure of classifier performance. In this work, the area under the ROC curve was 0.94.

Our algorithm was run on a Core i5, 2.53 GHz CPU and 4G of RAM. The segmentation time of a suspicious nodule candidate was 5.0 s.

**Conclusion**

In this paper, a new algorithm for lung nodule detection was proposed. At first, TM method was used for detection of suspicious nodule candidates. Suspicious nodule candidates included both nodules and blood vessels. The nodule candidate regions were segmented, and then blood vessels were removed from the candidate list. This algorithm was evaluated on a dataset of fifty thoracic CT scans.

Our method is efficient for accurate detection of lung nodules in CT images. We compared the detection rate with previous existing methods; the proposed method performs similar or better. Localized active contour model provided superiority of our algorithm. Furthermore, our algorithm decreased false-positive rate using a three-step procedure using shape features.

**Financial support and sponsorship**

None.

**Conflicts of interest**

There are no conflicts of interest.

**References**


