

## Intelligent Recognition of Lung Nodule Combining Rule-based and C-SVM Classifiers

Bin Li<sup>\*</sup>, Jing Zhang, Lianfang Tian, Li Tan, Shijie Xiang

*School of Automation Science and Engineering, South China University of Technology  
Guangzhou, Guangdong 510640, China*

*E-mail: binlee@scut.edu.cn*

*www.scut.edu.cn*

Shanxing Ou

*Department of Radiology, Guangzhou General Hospital of Guangzhou Command  
Guangzhou, Guangdong 510010, China*

Received 11 March 2011

Accepted 27 October 2011

### Abstract

Computer-aided detection(CAD) system for lung nodules plays the important role in the diagnosis of lung cancer. In this paper, an improved intelligent recognition method of lung nodule in HRCT combining rule-based and cost-sensitive support vector machine(C-SVM) classifiers is proposed for detecting both solid nodules and ground-glass opacity(GGO) nodules(part solid and nonsolid). This method consists of several steps. Firstly, segmentation of regions of interest(ROIs), including pulmonary parenchyma and lung nodule candidates, is a difficult task. On one side, the presence of noise lowers the visibility of low-contrast objects. On the other side, different types of nodules, including small nodules, nodules connecting to vasculature or other structures, part-solid or nonsolid nodules, are complex, noisy, weak edge or difficult to define the boundary. In order to overcome the difficulties of obvious boundary-leak and slow evolvement speed problem in segmentation of weak edge, an overall segmentation method is proposed, they are: the lung parenchyma is extracted based on threshold and morphologic segmentation method; the image denoising and enhancing is realized by nonlinear anisotropic diffusion filtering(NADF) method; candidate pulmonary nodules are segmented by the improved C-V level set method, in which the segmentation result of EM-based fuzzy threshold method is used as the initial contour of active contour model and a constrained energy term is added into the PDE of level set function. Then, lung nodules are classified by using the intelligent classifiers combining rules and C-SVM. Rule-based classification is first used to remove easily dismissible nonodule objects, then C-SVM classification are used to further classify nodule candidates and reduce the number of false positive(FP) objects. In order to increase the efficiency of SVM, an improved training method is used to train SVM, which uses the grid search method to search the optimal parameters of C-SVM and uses second order information to achieve fast convergence to select the Sequential Minimal Optimization(SMO) working set. Experimental results of recognition for lung nodules show desirable performances of the proposed method.

*Keywords:* Lung nodules, computer-aided detection, recognition, lung cancer, image segmentation, support vector machine

### 1. Introduction

Lung cancer is one of the most common malignant tumors. According to World Health Organization(WHO,

2004), Lung cancer, the most common cause of cancer-related death in men and women, is responsible for 1.3 million deaths worldwide annually, as of 2004. Early detection of lung cancer is critical for improving

---

<sup>\*</sup>The corresponding author, [binlee@scut.edu.cn](mailto:binlee@scut.edu.cn) or [binleemmboy@yahoo.com.cn](mailto:binleemmboy@yahoo.com.cn), the present address of the author is: School of Automation Science and Engineering, South China University of Technology, Wushan Road, Tianhe District, Guangzhou City, Guangdong Province, China, 510640

survival of this disease. Survival varies, depending on stage, overall health, and other factors, but the overall five-year survival rate for all persons diagnosed with lung cancer is 14%<sup>[1]</sup>. Guessous et al. <sup>[2]</sup> reported that the average 5-year survival rate of Europe lung cancer patients is 8%-12%, while the 5-year survival rate of the early stage patients can surpass 70%.

How to effectively diagnose lung cancer has become the subject of universal concern. Lung nodules in HRCT (high resolution computerized tomography) images are distinct signs of lung cancer. Lung nodules can be classified into two categories: solid nodules characterized by their high-contrast and ground-glass opacity (GGO) nodules with faint contrast and fuzzy margin<sup>[3]</sup>.

Radiologists will give detailed diagnosis results after reviewing carefully the CT images and testing the type (benign, malignant or cancerous) of nodules that are present. However, the interpretation of a great deal of CT images yielded by HRCT technique brings the huge workload to the radiologist, which will increase the false negative rate due to observational oversights. CAD systems help scan digital images, e.g. from computer tomography, for typical appearances and to highlight conspicuous sections, such as possible diseases. Therefore, Computer-aided detection/diagnosis (CAD) system for lung nodules plays the important role in the diagnosis of lung cancer<sup>[4]</sup>, which assists doctors in the interpretation of medical CT images and increases the detection of lung cancer by reducing the false negative rate due to observational oversights<sup>[3][5-8]</sup>.

CAD of lung nodules is fundamentally based on highly complex pattern recognition. In a CAD method of lung nodules, digital image data are prepared and analyzed in several steps:

(1) Image preprocessing, including reduction of artifacts, image noise reduction, etc.;

(2) Segmentation of ROIs (Regions of Interest), including differentiation of different structures in the image, e.g. heart, lung, ribcage, possible round lesions; In the CAD of lung nodules, segmentation of pulmonary parenchyma and nodule candidates are a necessary first step. Any computer system that analyzes the lungs and does not work on manually delineated regions of interest must incorporate an automatic lung segmentation. Armato and Sensakovic<sup>[9]</sup> illustrated the importance of accurate segmentation as a preprocessing step in a CAD scheme. In a nodule detection setting,

they showed that 5% - 17% of the lung nodules in their test data was missed due to the preprocessing segmentation, depending on whether or not the segmentation algorithm was adapted specifically to the nodule detection task. For finding pulmonary parenchyma and nodule candidates, so many techniques<sup>[8][10]</sup> have been reported: rule-based thresholding methods<sup>[9][11-15]</sup>, mathematical morphology<sup>[16]</sup>, genetic algorithm<sup>[16]</sup>, template matching of Gaussian spheres and discs<sup>[17]</sup>, clustering<sup>[8]</sup>, detection of (half) circles in thresholded images<sup>[18]</sup>, and filters enhancing (spherical) structures<sup>[7][19-20]</sup>.

Segmentation has been researched, and ongoing improvements in speed and quality of its techniques have broadened the field of applications, but has not reached the status of routine clinical use<sup>[4]</sup>. The original thorax image consists of background, lung tissue, blood vessel, trachea, bone, fat, muscle, and so on. But only the pulmonary parenchyma is useful indeedly for detection of lung nodules. So, the pulmonary parenchyma must be extracted before any other automatic segmentation and recognition of lesions. As the lung is essentially a bag of air in the body, it shows up as a dark region in CT scans. This contrast between lung and surrounding tissues forms the basis for the majority of the segmentation schemes. Segmentation of pulmonary parenchyma and different types of nodules is a difficult task. Reasons are as follows. On one side, thorax CT images, especially low-dose CT images are often affected by random noise arising in the image acquisition process and respiratory movements. The presence of noise not only produces undesirable visual quality but also lowers the visibility of low-contrast objects. On the other side, the excellent contrast between tissue and air on CT makes segmentation of an isolated solid nodule of reasonable size a simple task. But difficulties arise when a) the nodule is small; b) the nodule is connected to vasculature or other structures such as the pleura, fissures or abnormalities; c) the nodule is part-solid or nonsolid, in which case it can be typical weak edge and difficult to define the boundary; d) the data is noisy (typical for low-dose scans)<sup>[4]</sup>. Therefore, it is difficult to acquire the ideal segmentation effect only using general image-data-driven segmentation methods. For example, the edges extracted only by edge detection operator is usually discontinuous due to the blur and weak boundary in

medical image; or region growing and rule-based thresholding methods are not sensitive to small nodules and will fail to segment these pathological parts of the lung due to a lack of contrast between lung and surrounding tissues. Active contour models have been one of the most successful methods for image segmentation<sup>[21-24]</sup>. According to paper[24], there are several desirable advantages of active contour models over classical image segmentation methods, such as edge detection, thresholding, and region grow. First, active contour models can achieve sub-pixel accuracy of object boundaries. Second, active contour models can be easily formulated under a principled energy minimization framework, and allow incorporation of various prior knowledge, such as shape and intensity distribution, for robust image segmentation. Third, they can provide smooth and closed contours as segmentation results, which are necessary and can be readily used for further applications, such as shape analysis and recognition.

(3) Analyzing, selection and extraction of Features of ROIs;

Here, every detected ROI (Region of Interest) is analyzed individually for special characteristics, such as ① compactness, ② form, size and location, ③ reference to close-by structures / ROIs, ④ average greylevel value analyze within a ROI, ⑤ Proportion of greylevels to border of the structure inside the ROI, etc..

(4) Evaluation / classification of nodule candidates;

Here every ROI is evaluated and classified individually for the probability of a TP. If the detected structures have been classified as nodules, they are highlighted in the image for the radiologist.

There is a number of classification techniques used in the stage of the nodule detection CAD systems: rule-based or linear classifier<sup>[19][20][25]</sup>, Linear discriminant classifier (LDA)<sup>[26]</sup>, template matching<sup>[27]</sup>, nearest cluster<sup>[28]</sup>, neural network<sup>[29-30]</sup>, Bayesian classifier<sup>[31]</sup>, multilayer perception, radial basis function network (RBF), SVM (Support Vector Machine)<sup>[3]</sup>, and so on. SVM is a new developed learning algorithm based on SLT (Statistical Learning Theory), which is established by Vapnik in 1992<sup>[32]</sup>. SVM is based on the principle of structural risk minimization and VC theory to find the best compromise in the complexity and learning ability of the model according to limit sample information. SVM overcomes the problems of over-fitting, curse of dimensionality and local minimum in traditional

machine learning to a great extent; thereby it obtains better generalization ability<sup>[33-34]</sup>. As the result of the advantage of SVM, it has been widely applied in target recognition and image processing<sup>[3][35-36]</sup>.

Despite much effort being devoted to the computer-aided nodule detection problem, lung CAD systems remain an ongoing research topic<sup>[37]</sup>. One of the major difficulties is the detection of GGO nodules with low-dose thin-slice CT screening. Another two difficulties are the detection of nodules that are adjacent to vessels or the chest wall when they have very similar intensity; and the detection of nodules that are nonspherical in shape. In such cases, intensity thresholding or model-based methods might fail to identify those nodules.

In this paper, an improved lung nodule intelligent recognition method combining rule-based and SVM classifiers is proposed for detecting multiple nodules, including solid nodules and ground-glass opacity (GGO) nodules (part solid and nonsolid). The flowchart of the proposed recognition method for lung nodules in HRCT images is shown as Fig.1. The components of our algorithm utilize some existing methods in new ways. In particular, new efforts and some improvements have been made to adapt the methods into our recognition approach of lung nodules. The main contributions of the paper are summarized below:

1) A unified CAD framework for the detection of various types of solid nodules and ground-glass opacity (GGO) nodules (part solid and nonsolid).

2) In order to overcome the difficulties of obvious boundary-leak and slow evolution speed problem in segmentation of weak edge, an overall segmentation method is proposed in this paper, the image denoising and enhancing is realized by NADF method; candidate pulmonary nodules are segmented by the improved C-V level set method, in which the segmentation result of EM-based fuzzy threshold method is used as the initial contour of active contour model and a constrained energy term is added into the PDE of level set function.

3) Lung nodules are classified by using the intelligent classifiers based on rules and SVM (support vector machine). Rule-based classification in 3D space is first used to remove easily dismissible non-nodule objects, then cost-sensitive support vector machine (C-SVM) classification are used to further classify nodule candidates and reduce the number of false positive (FP) objects. In order to increase the efficiency of SVM, an improved training method is used to train SVM, which

uses the grid search method to search the optimal parameters of C-SVM and selects the SMO working set, using second order information to achieve fast convergence.

The proposed method has been trained and validated on a clinical dataset of 100 thoracic CT scans using a wide range of tube dose levels that contain 52 nodules determined by a ground truth reading process. The experimental results using the proposed method indicate the performances with sensitivity of 88.89% and specificity of 95.65%. Different types of challenging nodules such as small nodules, low-contrast part-solid and nonsolid nodules are identified.

## 2. Segmentation of ROIs

### 2.1. The description of segmentation process of ROIs

In this paper, pulmonary parenchyma and nodule candidates are segmented by using an overall segmentation method combining fuzzy thresholding, morphologic segmentation, NADF and improved geometric active contour model. The segmentation process of ROIs (pulmonary parenchyma and nodule candidates) is shown as Fig.2.

### 2.2. The description of segmentation process of ROIs

The original thorax image consists of background, lung tissue, blood vessel, trachea, bone, fat, muscle, and so on. But only the pulmonary parenchyma is useful indeedly for detection of lung nodules. So, the pulmonary parenchyma must be extracted before any other automatic segmentation and recognition of lesions.

As the lung is essentially a bag of air in the body, it shows up as a dark region in CT scans. This contrast between lung and surrounding tissues forms the basis for the majority of the segmentation schemes. So, in this paper the pulmonary parenchyma can be segmented by an overall segmentation method combining thresholding, edge dection, region growing, and morphology. In the segmentation process, the left and right pulmonary lobe, trachea and bronchia are firstly segmented, then the pulmonary parenchyma with smooth edge and no holes is extracted by using morphology method. The detailed steps are as follows:

- (1)The thorax is extracted.

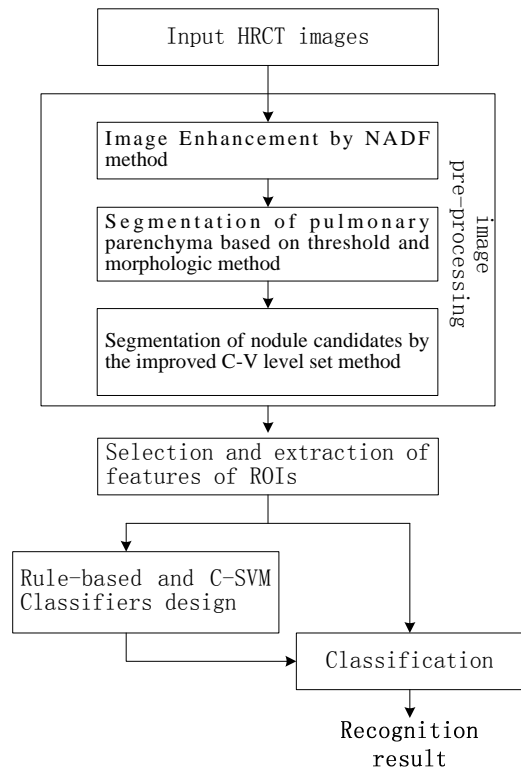


Fig. 1. Flowchart of the proposed recognition method for lung nodules

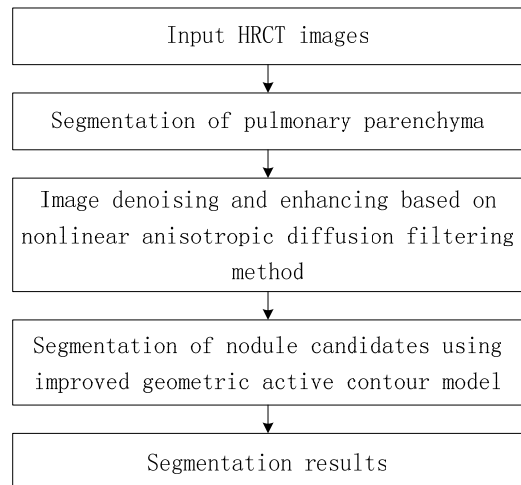


Fig. 2. Flowchart of segmentation process of ROIs

- ① Thorax is detected by using threshold segmentation and edge detection methods.
- ②All of edges in the binary image are detected by using edge detector.
- ③The contour of thorax is found by using contour track method.

④The thorax mask is acquired by using seed fill algorithm, and the primary pulmonary parenchyma is extracted by morphological processing method.

(2)Contours of left and right lung are extracted using contour track method, and their corresponding masks are filled up. If left and right lung is adjacent, they are divided by using the geometric character of lung and the topological saddle point in the junction region.

(3)The left and right mask are processed by morphological closing, respectively, and the irregular edge of mask is mended by using rolling ball method<sup>[42]</sup>, which is necessary to the detection of nodules that are adjacent to the chest wall.

(4)The final image of pulmonary parenchyma is segmented by using boolean operation AND between the original image and the mask of pulmonary parenchyma.

Results of each step are shown as Fig.3 and Fig.4.

**2.3. Lung image denoising and enhancing based on NADF method**

Thorax CT images, especially low-dose CT images are often affected by random noise arising in the image acquisition process and respiratory movements . The presence of noise not only produces undesirable visual quality but also lowers the visibility of low-contrast objects.

In order to recover the best estimate of the original image from its noisy version, and enhance the effect of ROIs, the image denoising and enhancing is realized by NADF method in this paper.

NADF is a technique aiming at reducing image noise without removing significant parts of the image content, typically edges, lines or other details that are important for the interpretation of the image<sup>[43]</sup>. It is described below.

The partial differential equation(PDE) which governs NADF is described as Eq.(1).

$$\frac{\partial u(x, y, t)}{\partial t} = \text{div}(g(|\nabla u_\sigma|^2)\nabla u) \quad (1)$$

Where  $u(x, y, t)$  is a filtered version of the original image;  $x, y$  are the horizontal ordinate value and the vertical ordinate value in a 2D image, respectively;  $t$  is a time variable;  $u_\sigma(x, y, t) = G_\sigma * u(x, y, t)$ ,  $G_\sigma$  is a 2D Gaussian distribution function of standard deviation  $\sigma$ , which denotes NAD is performed in

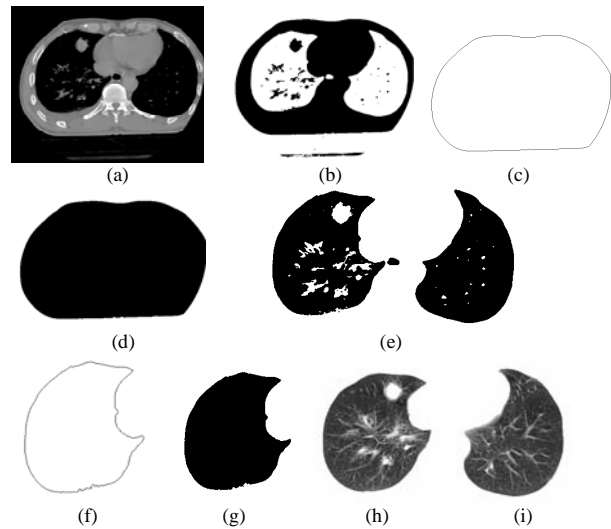


Fig. 3. Extraction processes of pulmonary parenchyma. (a)original image, (b)threshold segmentation of thorax, (c) contour track of thorax, (d)thorax mask, (e)threshold segmentation of pulmonary parenchyma, (f)contour of right lung, (g)mask of right lung, (h)right lung, (i)left lung

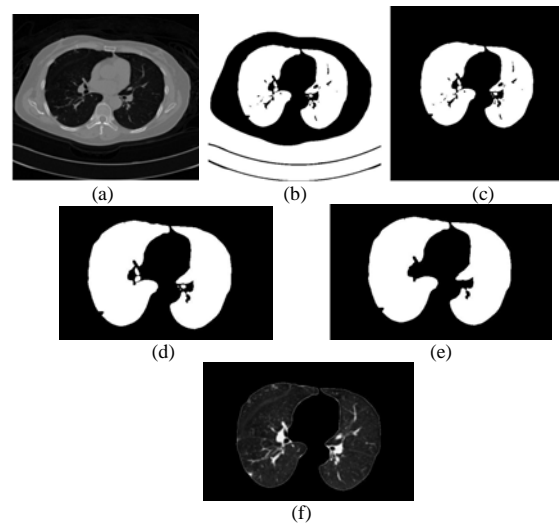


Fig. 4. Segmentation Results of Pulmonary Parenchyma. (a) original image, (b)optimal threshold segmentation, (c)primary segmentation result of Pulmonary Parenchyma, (d)filling up lung cavity, (e)eliminating trachea and bronchia, (f)mending image of pulmonary parenchyma

different scales;  $g(\bullet)$  is the monotonically degressive edge diffusion function, described as Eq.(2).

$$g(s) = \exp[-(\frac{s}{\lambda})^2] \quad (2)$$

Where  $\lambda$  is the diffusion constant.

The PDE Eq.(1) is ill conditioned, so it is regularized before solving the equation. An improved additional operator splitting(AOS) algorithm<sup>[43]</sup> with semi-implicit scheme is applied in this paper. The detailed implementation is described as follows.

**Step 1:** First, the image  $u^k$  at iteration  $k$  is convolved with a 2D Gaussian distribution function  $G_\sigma$  to obtain  $u_\sigma^k$ ; then the gradient norm  $|\nabla u_\sigma^k|$  of  $u_\sigma^k$  is computed, and the edge diffusion function  $g(|\nabla u_\sigma^k|)$  is solved.

**Step 2:** The image  $u(x, y, t)$  is filtered by 1D one-dimensional nonlinear diffusion filtering, respectively. A simple numerical scheme for solving this equation numerically is:

$$\begin{cases} u_1^{k+1} = (I - 2\tau A_x^k)^{-1} u^k \\ u_2^{k+1} = (I - 2\tau A_y^k)^{-1} u^k \end{cases} \quad (3)$$

Where,  $I$  is an identity matrix;  $\tau$  is the step length of time, and under the premise of absolute stability of the algorithm it is a trade-off between precision and efficiency;  $A_x^k, A_y^k$  are tridiagonal strictly diagonally dominant matrices, which are in connection with the edge diffusion function.  $u_1^{k+1}$  and  $u_2^{k+1}$  in Eq.(3) is solved using the Thomas algorithm.

**Step 3:**  $u^{k+1} = \frac{1}{2}(u_1^{k+1} + u_2^{k+1})$ ,  $u^{k+1}$  is the result after the current iteration.

**Step 4:** Repeat step(1)-(3), until the termination condition of iteration is satisfied.

ROIs are enhanced by using NADF method based on AOS algorithm<sup>[43]</sup>, which can remove some discontinuous and fine tissue and blood in pulmonary parenchyma. Moreover, nodules with different size are denoised and enhanced by changing the coefficient of variation or mask size of Gaussian model.

The filtering result of iteration 2 using NAD filtering with  $\lambda = 10$  and  $\tau = 5$  is shown as Fig.5.

#### 2.4. Segmentation of ROIs based on improved geometric active contour model

It is difficult to acquire the ideal segmentation effect only using general image- data-driven segmentation methods. On one side, thorax CT images, especially low-dose CT images are often affected by random noise arising in the image acquisition process and respiratory

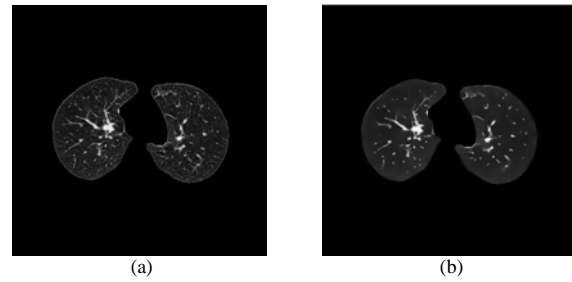


Fig. 5. The filtering result of pulmonary parenchyma using nonlinear anisotropic diffusion filtering. (a)segmentation of lung, (b)the filtering result.

movements. The presence of noise not only produces undesirable visual quality but also lowers the visibility of low-contrast objects. On the other side, the excellent contrast between tissue and air on CT makes segmentation of an isolated solid nodule of reasonable size a simple task. But difficulties arise when a) the nodule is small; b) the nodule is connected to vasculature or other structures such as the pleura, fissures or abnormalities; c) the nodule is part-solid or nonsolid, in which case it can be typical weak edge and difficult to define the boundary; d)the data is noisy (typical for low-dose scans).

Active contour models have been one of the most successful methods for image segmentation. Yet boundary leakage and slow evolvement speed problem often occur when the active contour model is applied to segment weak edge objects such as lung nodules. In order to overcome the difficulties of obvious boundary-leak and slow evolvement speed problem in segmentation of weak edge, pulmonary nodule candidates are segmented by improved C-V level set method in this paper. In this paper, the segmentation result of EM-based fuzzy threshold method is used as the initial contour of active contour model, and a constrained energy term is added into the PDE of level set function  $\phi$ , the modification make the active contour model more robust to solve boundary-leak problem, moreover, its segmentation contours are more smooth.

##### 2.4.1. EM-based fuzzy threshold Segmentation of ROIs

Lung nodules are related to statistical characteristics of intensity in CT images because of CT image principle. For example, lung nodules are often calcified tissue, whose intensity and that of gasbag are high-contrast. So

in this paper, the initial ROIs are segmented by using EM-based threshold method. First, the probability that each pixel belongs to each class is computed by using EM algorithm; then ROIs are segmented by using threshold method.

A set  $\{x_1, x_2, \dots, x_n\}$  of intensity values of  $N$  pixels in the image is represented as a observation sample, and its intensity value data are independent and identically.

Suppose a image consists of  $M$   $(1, 2, \dots, k, \dots)$  non-intersect region  $\omega_k$ , and all of intensity values of each region obey the Gaussian distribution with mean  $\mu_k$  and standard deviation  $\sigma_k$ . Then the distribution of intensity values in the whole image is described as a Gaussian mixture distribution, and its probability density function is:

$$f(x|\theta) = \sum_{j=1}^M \alpha_j f_j(x|\theta_j) \quad (4)$$

Then the distribution of intensity values in the whole image is described as a Gaussian mixture where  $f_k(x|\theta_k)$  is the density function of the  $k$  class Gaussian distribution,  $\theta_k$  is its parameter vector, including mean  $\mu_k$  and covariance matrix  $\Sigma_k$ ;  $\alpha_i$  is prior probability,  $\sum \alpha_j = 1$ .

Therefore, the logarithm likelihood function of image intensity distribution is:

$$L(x|\theta) = \sum_{i=1}^N \log(\sum_{j=1}^M \alpha_j f_j(x_i|\theta_j)) \quad (5)$$

$\theta$  of Eq.(5) is solved by EM algorithm. Its implementation is described as follows.

The first step(the E-step) of the EM algorithm is to compute the expectation of the image complete-data log-likelihood.

$$E^{(t)}[L(x|\theta)] = \sum_{i=1}^N \sum_{j=1}^M \log \alpha_j p^{(t)}(j|x_i, \theta) + \sum_{i=1}^N \sum_{j=1}^M \log f_j(x_i|\theta_j) p^{(t)}(j|x_i, \theta) \quad (6)$$

where  $p^{(t)}(j|x_i, \theta)$  is the conditional probability of each pixel in each class, which is computed by Bayesian formula shown as Eq.(7).

$$p^{(t)}(j|x_i, \theta) = \frac{\alpha_j f_j(x_i|\theta_j)}{\sum_{k=1}^M \alpha_k f_k(x_i|\theta_k)} \quad (7)$$

Where  $p^{(t)}(j|x_i, \theta)$  is the result at iteration  $t$ .

The second step(the M-step) of the EM algorithm is to maximize the expectation computed in the first step by using maximum likelihood estimation method.

$$\alpha_j^{(t+1)} = \frac{1}{N} \sum_{j=1}^N p^{(t)}(j|x_i, \theta) \quad (8)$$

$$\mu_j^{(t+1)} = \frac{\sum_{i=1}^N x_i p^{(t)}(j|x_i, \theta)}{\sum_{i=1}^N p^{(t)}(j|x_i, \theta)} \quad (9)$$

$$\Sigma_j^{(t+1)} = \frac{\sum_{i=1}^N p^{(t)}(j|x_i, \theta)(x_i - \mu_j^{(t+1)})(x_i - \mu_j^{(t+1)})^T}{\sum_{i=1}^N p^{(t)}(j|x_i, \theta)} \quad (10)$$

Repeat the above two steps, until the termination condition of iteration is satisfied.

Finally, ROIs are segmented by using thresholds computed by the aforementioned EM algorithm.

#### 2.4.2. Segmentation of ROIs based on improved C-V level set method

As mentioned above, it is difficult to acquire the ideal segmentation effect only using general image-data-driven segmentation methods. In order to overcome the difficulties of obvious boundary-leak and slow evolvement speed problem in segmentation of weak edge, pulmonary nodule candidates are segmented by improved C-V level set method in this paper.

The traditional C-V model is described below.

The image  $u$  is divided into two regions by the evolving curve  $C$ : the object region  $\Omega_1$  (inside( $C$ )) and background region  $\Omega_2$  (outside( $C$ )). The C-V energy function model<sup>[30][33]</sup> is:

$$E(C, c_1, c_2) = \mu \oint_C ds + \lambda_1 \iint_{\Omega_1} |u(x, y) - c_1|^2 dx dy + \lambda_2 \iint_{\Omega_2} |u(x, y) - c_2|^2 dx dy \quad (11)$$

Where,  $C$  is any other variable curve;  $c_1, c_2$  depending on  $C$ , are the averages of image intensity inside  $C$  and outside  $C$ , respectively; the first term on the right hand side of the equation is the length of

evolving curve  $C$ , and the second and the third term are the internal energy and exterior energy, respectively;  $\mu, \lambda_1, \lambda_2$  are weights of the corresponding term, respectively;  $u(x, y)$  is the intensity value in  $(x, y)$ . The sum of two energy terms is minimized if  $C$  is the boundary.

The C-V model is solved by using variational level set approach, and its partial differential equations(PDE) defined by level set function  $\phi$  is:

$$\frac{\partial \phi}{\partial t} = \delta_\epsilon(\phi) [\mu \operatorname{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) - \nu - \lambda_1 (I_0 - c_1)^2 + \lambda_2 (I_0 - c_2)^2] = 0 \tag{12}$$

where,  $\delta_\epsilon(\phi)$  is slightly regularized versions of *Dirac* measure  $\delta(\phi)$ ;  $\mu, \nu, \lambda_1, \lambda_2$  represents the weight of the corresponding energy term, respectively;  $I_0$  is the object region;  $c_1, c_2$  is the average intensity value inside and outside contour.

It is difficult to acquire the ideal segmentation effect to segment weak edge of medial image only using the traditional C-V model segmentation methods. An example is shown as Fig.6.

Because the traditional C-V model is evolved based on gradient and intensity information, obvious boundary-leak and slow evolvement speed problem in segmentation of lung nodule candidates only using the traditional C-V model arise when the nodule is weak edge or the data is noisy (typical for low-dose scans). In order to overcome the difficulties of obvious boundary-leak and slow evolvement speed problem in segmentation of weak edge, pulmonary nodule candidates are segmented by improved C-V level set method in this paper. First, according to the overall and local region information the intensity value of each pixel in some region is unified to average value of its corresponding region, which will improve the condition of intensity inhomogeneity in the same region; and the segmentation result of EM-based fuzzy threshold method is used as the initial contour of active contour model. Then, a constrained energy term is added into the PDE of level set function  $\phi$ ; the parameter  $\nu$  is

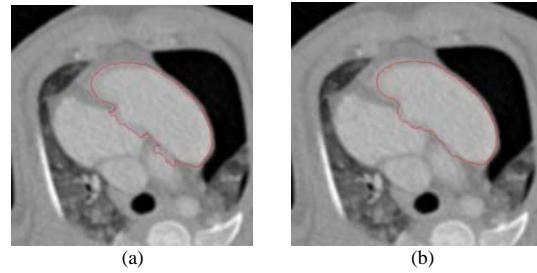


Fig. 6. Illustration of difficulties of traditional CV segmentation method. (a) boundary-leak, (b) incomplete convergence

added into the original level set function  $\phi_0$ , that is: the level set function  $\phi = \nu \phi_0$ , here  $\nu$  is the distance weight,  $\nu > 1$ . The modification make the active contour model more robust to solve boundary-leak problem, moreover, its segmentation contours are more smooth. So it is fit for the segmentation of blur edge and avoids expensive re-initialization of the evolving level set function.

#### 2.4.3. Segmentation results of ROIs

Experimental results indicate that the proposed method is satisfactory, which are separately shown in Figs 7-9. Fig.7, Fig.8 and Fig.9 show some segmentation results of typical lung nodules, including GGO(ground-glass opacity), solitary and adhesive lung nodules. Fig.7(b), Fig.8(b) and Fig.9(b) are the denoised and enhanced results using the proposed method, in which the edge information has been enhanced and some discontinuous and fine tissue and blood have been removed. Fig.6(c), Fig.7(c) and Fig.8(c) is the segmentation results using the proposed method, in which, some unwanted and meaningless information are removed. In Fig.6(c), Fig.7(c) and Fig.8(c), the red pane and arrows denoted nodules which are recognized by the following recognition method. Experimental results show that the segmentation contours of nodules are segmented completely and smoothly.



### 3. Feature selection and extraction and rule-based classifier

How to select and extract features is the key for intelligent detection of lung nodule. According to the clinical manifestation and pathological features of lung nodules, some useful intensity, shape and texture features are selected and extracted from diagnosed samples and disturbed samples. Then a rule-based classifier is designed to remove easily dismissible non-nodule objects.

#### 3.1. Clinical manifestation and pathological features of lung nodules in CT

Clinical manifestation and pathological features of lung nodules in CT include position, shape, lobulation, burr sign, notch, cavitation, vacuole sign, halo sign, calcification, vessel convergence sign, satellite lesion, pleural retraction sign, etc.

#### 3.2. Feature extraction of lung nodules

Feature selection and extraction play an important role on the intelligent recognition of lung nodules. In this paper, typical features of lung nodules include intensity, shape and texture features.

##### (1) Intensity features

Intensity features are basic and simple features of lung nodules, including histogram, gray-level, intensity mean value  $\bar{f}$  and variance  $\sigma_f^2$  of ROI.

##### (2) Shape features

Most of signs of clinical manifestation and pathological features of lung nodules are from shape features, so it is important to select and extract shape features. Shape features consist of global shape features, which include perimeter, area, diameter, centroid, ellipticity, circularity, slenderness, rectangle degree etc., and local shape features which include inflexion of contour, important feature points, etc.

##### 1) Invariant moments

The p+q order origin moment  $m_{pq}$  of a  $M \times N$  image is defined by Eq.(13).

$$m_{pq} = \sum_i \sum_j i^p j^q f(i, j) \quad (13)$$

Where  $f(i, j)$  is the intensity value of  $(i, j)$ .

The p+q order central moment  $\mu_{pq}$  is defined by Eq.(14).

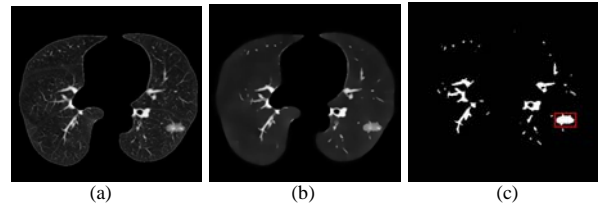


Fig. 7. Result of the suspected ground-glass opacity lung nodules. (a)original image, (b)NAD filtering enhancing, (c)the proposed method.

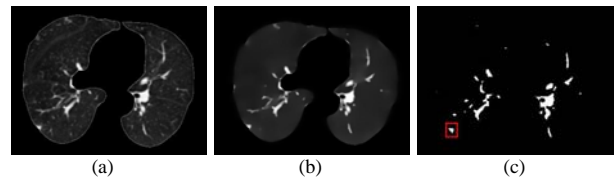


Fig. 8. Result of the suspected adhesive lung nodules segmentation. (a)original image, (b)NAD filtering enhancing, (c)the proposed method.

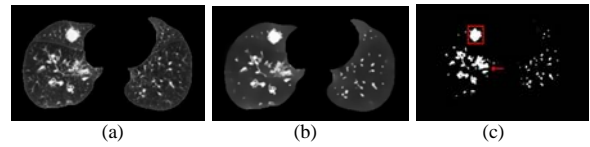


Fig. 9. Result of the suspected solitary lung nodules segmentation. (a)original image, (b)NAD filtering enhancing, (c)the proposed method.

$$\mu_{pq} = \sum_i \sum_j (i - i_0)^p (j - j_0)^q f(i, j) \quad (14)$$

Where  $i_0 = m_{10} / m_{00}$ ,  $j_0 = m_{01} / m_{00}$ ,  $(i_0, j_0)$  is the centroid.

①The centroid  $(i_0, j_0)$

$$i_0 = m_{10} / m_{00}, \quad j_0 = m_{01} / m_{00} \quad (15)$$

②Long axis and short axis

The long axis and short axis are shown as Eq.(16) and Eq.(17), respectively.

$$a = 2 \times [2(\mu_{20} + \mu_{02} + \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}) / \mu_{00}]^{1/2} \quad (16)$$

$$b = 2 \times [2(\mu_{20} + \mu_{02} - \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}) / \mu_{00}]^{1/2} \quad (17)$$

Where,  $\mu_{pq}$  is the p+q order central moment. The long axis is used to measure the diameter of the lung nodule.

③Flattening of region is:

$$e = a / b \quad (18)$$

The closer to circle the shape of region is, the closer to 1  $e$  is.

2)geometric meature

① Perimether  $P$  is

$$P = A_{odd} + \sqrt{2}A_{even} \quad (19)$$

Where  $A_{odd}$  is sum of odd number in chain code,  $A_{even}$  is sum of even number in chain code. 8-connected chain code is used in this paper.

②Area  $A$  is:

$$A = \text{total of pixels in some region} \quad (20)$$

③Circularity is:

$$Cir = r_i / r_c \quad (21)$$

Where  $r_i$  is radius of inscribed circle,  $r_c$  is radius of circumscribed circle.

④Rectangularity  $R$  is:

$$R = A / (H \times W) \quad (22)$$

Where  $W$  is width,  $H$  is height. Rectangularity reflects departure degree of region and rectangle.  $R \in [0,1]$ . The closer to rectangle the shape is, the closer to 1  $R$  is.

⑤Slenderness is:

$$S = \frac{\min(W, H)}{\max(W, H)} \quad (23)$$

⑥Compact ratio/ compactness is:

$$F = 4\pi A / P^2 \quad (24)$$

Where,  $P$  is perimeter of region contour,  $A$  is area of region. Compact ratio reflects degree closing to circle and smoothness of region. And compact ratio can detect irregular feature of edge. The rougher and flatter edge is, the smaller  $F$  is.

⑦Concavity ratio is:

$$E = S_e / S \quad (25)$$

Where  $S$  is area of concave region,  $S_e$  is difference of convex hull and original region. Concavity ratio reflects cupped degree of region boundary.

(3)Texture features

It is necessary to explain and describe the appearance of tissue in detection of medical image, such as smoothness, granularity, regulation, uniformity, and so on. All of these characters are related to local intensity change of image, which can be quantified by texture measurement. In this paper, co-occurrence matrix is used to quantified texture measurement. Let

the gray level of image be  $L$ , distance  $\delta = 2$  and four directions,  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ , are considered here. Four parameters of co-occurrence matrix are as follows:

Energy  $f_1$  is uniformity measurement of gray distribution:

$$f_1 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P_\delta^2(i, j) \quad (26)$$

Where,  $P_\delta(i, j)$  is normalized co-occurrence matrix.

Entropy  $f_2$  is:

$$f_2 = -\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P_\delta(i, j) \log P_\delta(i, j) \quad (27)$$

Correlation  $f_3$  is:

$$f_3 = \frac{\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i - \mu_1)(j - \mu_2) P_\delta(i, j)}{\sigma_1^2 \sigma_2^2} \quad (28)$$

Where,  $\mu_1 = \sum_{i=0}^{L-1} i \sum_{j=0}^{L-1} P_\delta(i, j)$ ,  $\mu_2 = \sum_{j=0}^{L-1} j \sum_{i=0}^{L-1} P_\delta(i, j)$ ,

$$\sigma_1^2 = \sum_{i=0}^{L-1} (i - \mu_1)^2 \sum_{j=0}^{L-1} P_\delta(i, j)$$

Inertia moments  $f_4$  is:

$$f_4 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i - j)^2 P_\delta(i, j) \quad (29)$$

Eigenvalue  $f_{uc}$  of texture matrix of four parameters for one region is:

$$f_{uc} = a_1 f_1 + a_2 f_2 + a_3 f_3 + a_4 f_4 \quad (30)$$

Where,  $a_i$  is weight coefficient,  $a_1 = 0.45$ ,  $a_2 = 0.40$ ,  $a_3 = 0.09$ ,  $a_4 = 0.06$ .

3.3. Feature selection of lung nodules in CT

According to pathomorphology of lung nodule lesions, a small lesion show circle because it have not encountered many obstacles in its growing process; when the lesion grows to a certain degree, it shows quasi-circular, and lobulation sign and notch sign appear due to the barrier of neighborhood bronchus trees and blood vessels; as the lesion grows further, it will oppress the bronchia, blood vessel or pleura and shows irregular shape, and continuous burr sign and lobulation will appear. Therefore, shape features are still the most important evidence for detection of true lung nodules. Moreover, burr sign, lobulation and cavitation with GGO edge show blur and uneven, which are the most important pathomorphological features in mid-later growing season and can be represented by compact ratio.

Furthermore, nodule candidate extraction is often disturbed by bronchus and blood vessels in pulmonary hilar.

Therefore, according to expert knowledge and experimental results, shape features of typical samples, such as concavity ratio, circularity, flattening and compact ratio are summarized in this paper.

Moreover, the blood vessel whose transect shows quasi-circular is easily mistaken for true positive nodules, shown as Fig.10. According to the medical knowledge<sup>[44]</sup>, main differences between lung nodules and blood vessels are as follows:

①When the blood vessel is parallel to the slice, the lung nodule shows circular or quasi-circular, yet the blood vessel shows rectangular;

②When the blood vessel is vertical to the slice, the blood vessel and lung nodule are both circular, yet CT value of the blood vessel is higher than that of nodule with the same size; moreover, the position of the blood vessel almost keeps unchanged;

③CT values of lung nodule obey the Gaussian distribution, yet the intensity values of transect of blood vessel are almost uniform. From the statistical result of samples, intensity values of interior of lung nodule are  $-100 \sim 100$ , and their variance values are  $25 \sim 60$ ; intensity values of blood vessel are often less than  $-100$ , and their variance value are less than 25.

### 3.4. Rule-based classifier

In this paper, a rule-based classifier is designed to remove easily dismissible non-nodule objects from lung nodule candidates. That is, features of lung nodule candidates are extracted in section 3.3 and integrated into the corresponding feature matrices, and lung nodule candidates are classified by the cascaded decision-making method combining with rules and multi-feature tracking. First, lung nodule candidates are classified by the classification rules based on shape, intensity and texture features; then corresponding features of nodule candidates in multiple CT images are tracked and fused as needed, and the nodule candidates are detected further.

A lung nodule often has the following obvious characteristics: (1) Its position is relatively unchanged; (2) Its area has a great deal of change; (3) It shows oval-like and changes little, etc.. So the classification method based on multi-feature tracking is proposed, which is fit for CT sequence images. Here, the concept “integral”

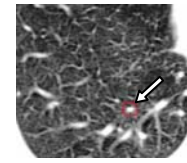


Fig. 10. The blood vessel similar to nodules.

$I_{sum}$  is borrowed to reflect the change degree of the nodule candidate in multi-feature tracking. And rules are as follows.

(1) The position change between the nodule candidate in  $i$ -th CT image and another nodule candidates in  $i+1$ -th CT image is smaller than the threshold, then they are looked as the same nodule.

(2) The position, area, compact ratio and intensity change between the nodule candidate in  $i$ -th CT image and another nodule candidates in  $i+1$ -th CT image are smaller than the threshold, respectively, then  $I_{sum} + 1$  and the feature is fused.

(3) When the selected CT images are processed, the nodule candidate is classified as a nodule if the corresponding “integral” value is greater than the threshold.

In this paper, classification rules based on shape, intensity and texture features are as follows:

① If concavity ratio  $E > 0.6$  and circularity  $Cir < 0.5$ , a false positive nodule, belonging to the blood vessel or bronchial.

② If concavity ratio  $0.15 < E < 0.6$ , and flattening  $e > 1.75$ , or circularity  $Cir < 0.5$ , it is a false positive nodule, belonging to the blood vessel, bronchial or value tissue.

③ If concavity ratio  $E < 0.15$ , and flattening  $e > 1.75$ , it is a false positive nodule, belonging to the blood vessel, bronchial or value tissue.

④ If concavity ratio  $E < 0.15$ , circularity  $Cir > 0.8$ , compact ratio  $F > 0.7$ , and flattening  $e < 1.5$ , it is a true positive nodule in early stage, which shows quasi-circular and whose contour is smooth.

⑤ If concavity ratio  $E < 0.6$ , circularity  $0.6 < Cir < 0.8$ , flattening  $e < 1.5$ , and compact ratio  $0.5 < F < 0.7$ , it is a true positive nodule in early-mid stage, whose circularity descends and whose edge shows lobulation sign.

⑥ If concavity ratio  $0.15 < E < 0.6$ , circularity  $0.5 < Cir < 0.7$ , flattening  $e < 1.5$ , and compact ratio  $F < 0.5$ , it is a true positive nodule in terminal stage,

which shows shows burr sign, moderate or deep lobulation sign. Moreover its edge gets blur.

After the above identification, filtered nodule candidates are classified into lung nodules or blood vessel. The rules are as follows:

⑦ Intensity values of interior of lung nodule are  $-100 < \bar{f} < 100$ , and their variance values are  $25 < \sigma_f^2 < 60$ ; yet intensity values of the blood vessel are less than -100, and their variance are less than 25.

⑧ If meeting rule⑦, and its texture feature  $f_{uc}$  is bigger than threshold 0.50, it is classified into the lung nodule.

#### 4. Classification of lung nodules based on rules and SVM(support vector machine) classifier

As mentioned above, lung nodules are classified by using the intelligent classifiers based on rules and SVM(support vector machine). First, rule-based classification is used to remove easily dismissible nonnodule objects. Then, cost-sensitive support vector machine (C-SVM) classification are used to further classify nodule candidates and reduce the number of false positive(FP) objects. Here, the nodule candidates filtered by the rule-based classification are used as the training and testing samples set of C-SVM, and their intensity, shape and texture features are used as the input parameters of C-SVM.

The C-SVM is used to solve the problem of un-balanced dataset. The nodule candidates set is a typical un-balanced dataset, in which the nodule samples are far less than the nonnodule samples, so the C-SVM is fit for the classification of samples in an un-balanced dataset.

In this paper, nodule candidates are regarded as the samples set  $X = [x_1, x_2, \dots, x_N]^T$ , where  $N$  denotes the number of total samples, and  $x_i (i = 1, 2, \dots, N)$  is a row vector, representing the features set of any sample.

Here, 13 features are taken as the inputs of C-SVM, which include the area, diameter, perimeter, rectangle degree, flattening, circularity, slenderness, intensity mean value, variance, energy, compactness, concavity ratio, eigenvalue, according to the features selected and extracted in section 3. The input data of C-SVM are normalized to  $[0, 1]$ , and the class label of C-SVM  $y_i \in \{+1, -1\}$ ,  $i = 1, 2, \dots, N$  (where +1 corresponding to a nodule and -1 to a nonnodule) of the corresponding sample  $x_i$  is also given. Then the samples label set is  $Y = [y_1, y_2, \dots, y_N]^T$  and the dataset can be

represented as  $(X, Y)$ . The decision function of the C-SVM is given as<sup>[51-52]</sup>:

$$f(x) = \text{sgn}[\sum_{i=1}^l a_i y_i K(x_i, x_j) + b] \quad (31)$$

Where  $K(x_i, x_j)$  is a nonlinear kernel function,  $b \in R$  and  $a$  is constrained as follows:  $0 \leq a_i \leq C_+$ , for  $y_i = +1$ , and  $0 \leq a_i \leq C_-$ , for  $y_i = -1$ . Where  $C_+$  and  $C_-$  are penalties for class +1 and -1, respectively.

The SVM requires the solution of the following optimization problem in order to seek the best hyperplane of  $\omega \bullet x + b = 0$  ( $\omega$  is the normal vector of the optimal hyper plane,  $b$  is the bias), so the formulation of the quadratic program of the C-SVM classifier is shown as Eq.(32).

$$\begin{aligned} \min & \frac{1}{2} \|\omega\|^2 + C_+ \sum_{y_i=+1} \xi_i + C_- \sum_{y_i=-1} \xi_i \\ \text{s.t.} & \forall i, y_i(\omega \bullet x_i + b) + \xi_i \geq 1 \\ & \forall i, \xi_i \geq 0 \end{aligned} \quad (32)$$

Where  $\xi_i$  is the slack variable.

Optimization problems of Eq.(32) can be converted to their dual form as Eq.(33) by differentiating the Lagrange multiplier.

$$\begin{aligned} \max & \sum_{i=1}^l a_i - \frac{1}{2} \sum_{i,j=1}^l a_i a_j y_i y_j K(x_i, x_j) \\ \text{s.t.} & \sum_{i=1}^l a_i y_i = 0 \\ & 0 \leq a_i \leq C_+, y_i = 1 \\ & 0 \leq a_i \leq C_-, y_i = -1 \end{aligned} \quad (33)$$

Where  $a_i$  is the Lagrange multiplier of  $x_i$ .

According to papers[38][39], the training algorithms of the traditional C-SVM<sup>[40]</sup> are slow, especially for large data problems. In this paper, in order to increase the efficiency of SVM, an improved training method is used to train SVM, which uses the grid search method to search the optimal parameters of C-SVM and selects the Sequential Minimal Optimization (SMO) working set, using second order information to achieve fast convergence.

In this proposed method, the penalties  $C_+$  and  $C_-$  are defined as Eq.(34):

$$\frac{C_+}{C_-} = \frac{n_-}{n_+} \quad (34)$$

where  $n_+$  and  $n_-$  are the number of positive samples and negative samples, respectively.

The RBF kernel function is used in this paper, defined as Eq.(35):

$$K(x_1, x_2) = \exp(-\gamma \|x_1 - x_2\|^2) \quad (35)$$

Where  $\gamma$  is the parameter.

In this paper, nodule candidatas are classified by k-Cross Validdation training and testing SVM. And the grid search method is used to search the optimal parameters  $C_$  and  $\gamma$  of C-SVM in this paper. In the grid search method, there should be a criterion to determine the optimal parameters. However, how to find a best criterion is still a difficult problem in the case of un-balanced dataset. Sensitivity and specificity are often used to measure the performance of the classification system; but they are often a tradeoff. The criterion of AUC<sup>[45]</sup> is chosen to train C-SVM. When AUC reaches the maximum, the minimums of  $C_$  and  $\gamma$  meeting the AUC condition will be chosen.

## 5. Experimental Results

A database of 100 thoracix CT scans from Guangzhou General Hospital of Guangzhou Command, a comprehensive First Class of the third Grade Hospital in Guangzhou, was used to evaluate the effectiveness of the proposed method. The medical CT slices used are data set with an intensity value of 16bits and the resolution of 512\*512 .

50 slices are chosen as training data set, and other 50 slices are chosen as test data set in the paper. Lung nodules in CT images are solid or GGO(part solid or nonsolid), whose sizes are from 4mm to 30mm. Moreover, locations of nodules are uncertain, some are isolated, others are adhered to lung wall or blood vessels. There are 254 candidate ROIs, including 52 (lung nodules) and 204 negative samples (non-nodules), shown as Table 1.

Table 1. Experiment data sets

Data set	Lung-nodule
positive samples	52
negative samples	204
feature number	13

Some detected results of different types of nodules are showed in Fig.11-Fig.14. Fig.11 (a) is a CT slice with a 8mm solid nodule and an adhesive nodule which is adjacent to lung wall; Fig.12(a) is a CT slice with a

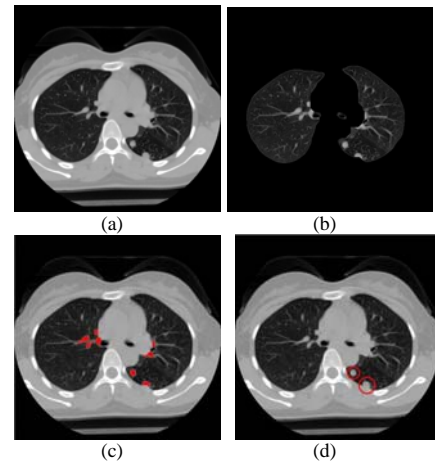


Fig. 11. Recognition results of the case containing a solitary nodule and an adhesive nodule. (a)CT original image, (b) Lung parenchyma segmentation, (c) ROIs extraction, (d) classification Results.

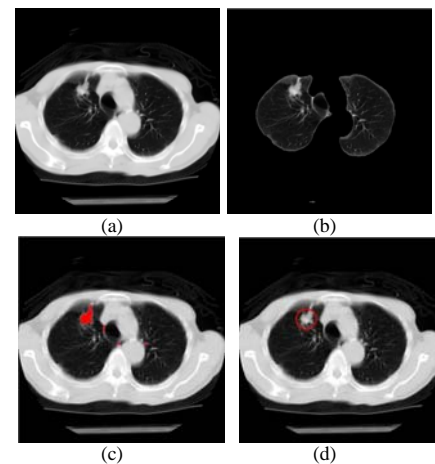


Fig. 12. Recognition results of the case containing an adhesive nodule. (a) CT original image, (b) Lung parenchyma segmentation, (c) ROIs extraction, (d)Classification Results.

nodule which is adjacent to the blood vessel; Fig.13(a) is a CT slice with a GGO nodule; and Fig.14 (a) is a CT slice including a micro-nodule with the diameter 4mm. Fig.11(b), Fig.12(b), Fig.13(b) and Fig.14(b) are the corresponding segmentation results of lung parenchyma, respectively. Fig.11(c), Fig.12(c), Fig.13(c) and Fig.14(c) are the corresponding ROIs extraction results (marked in red regions), respectively; Fig.11(d), Fig.12(d), Fig.13(d) and Fig.14(d) are the corresponding recognition results of the system (marked in red circle), respectively.

Experimental results shown the classification results of the proposed method are satisfactory. The classification results of experimental data set Table 1 are classified by using the proposed method are shown as Table 2, and ROC curves are given in Fig.15.

In this paper, K-fold CV was performed to train and test C-SVM classifier, and different results were obtained using different K values. In the grid search method, define  $\gamma = bestg$ ,  $C_ = bestc$ . In Table 3, when sensitivity reaches 88.89%, specificity is 95.65%(K=8), and the performance of the SVM is considered the best.

Comparing with different CAD models covered in paper<sup>[3-4]</sup>, it seems that the proposed method's relatively high detection rate, fast computation, and applicability to different imaging conditions and nodule types shows much promise for clinical applications. As a rule, nodule detection systems consist of several steps: a) preprocessing; b) candidate detection; c) false positive reduction; d) classification. Most often one tries to cheaply segment and detect the potential nodules and drastically reduce the number of these FPs. They are necessary to improve the detection rate and reduce the number of false positives. The reason why the proposed method has a better performance for detecting both solid nodules and ground-glass opacity nodules are as follows.

1)candidate detection; The purpose of candidate detection is to avoid missing potential nodules. Moreover in this step, the precise segmentation of potential nodule is often a necessary step to computer analysis, extraction and computation of features, which is important for false positive reduction. As mention before, our solution to efficiently segment the potential nodule objects involves three steps: i) the lung parenchyma is extracted based on threshold and morphologic segmentation method; ii) the image denoising and enhancing is realized by NADF method; iii) candidate pulmonary nodules are segmented by the improved C-V level set method. The proposed overall segmentation method will efficiently overcome the difficulties caused noise and weak edge.

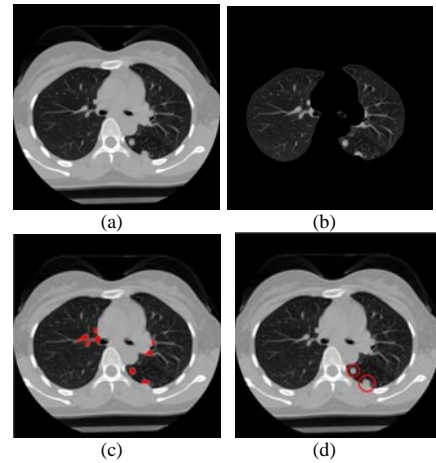


Fig. 13. Recognition results of the case containing a ground-glass opacity nodule. (a) CT original image, (b) Lung parenchyma segmentation, (c) ROIs extraction, (d) Classification results.

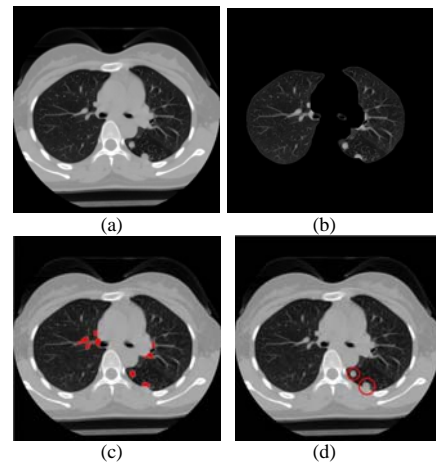


Fig. 14. Recognition results of the case including a micro-nodule. (a)CT original image, (b) Lung parenchyma segmentation, (c)ROIs extraction, (d)Classification Results.

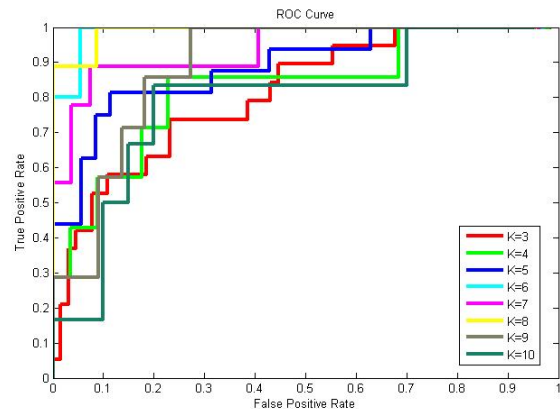


Fig. 15. ROC curves using the grid search for optimal parameters of C-SVM classifier for different K values.

Table 2. K-fold CV results using the grid search for optimal parameters of C-SVM classifier.

Parameters	TN	FN	TN	FP	Sensitivity	Specificity	AUC
K=3,bestc=0.5,bestg=1	13	6	50	15	0.6842	0.7692	0.7960
K=4,bestc=4,bestg=0.25	7	0	0	57	1.0	0.0	0.8095
K=5,bestc=4,bestg=2	16	0	0	57	1.0	0.0	0.8571
K=6,bestc=2,bestg=1	5	0	0	37	0.0	1.0	0.9622
K=7,bestc=2,bestg=4	0	9	27	0	0.0	1.0	0.9012
K=8,bestc=2,bestg=8	8	1	22	1	0.8889	0.9565	0.9469
K=9,bestc=0.5,bestg=0.125	7	0	0	22	1.0	0.0	0.8442
K=10,bestc=8,bestg=0.5	6	0	0	20	1.0	0.0	0.7417

2) false positive reduction; In this paper, Lung nodules are classified by using the intelligent classifiers based on rules and SVM(support vector machine). First, rule-based piecewise linear classifier in 3D space is used to remove easily dismissible nonodule objects. Then, cost-sensitive support vector machine (C-SVM) classification are used to further classify nodule candidates and reduce the number of false positive(FP) objects. Moreover, an improved training method is used to train SVM, which uses the grid search method to search the optimal parameters of C-SVM.

## 6. Conclusions and Discussion

In this paper, an improved intelligent recognition method of lung nodule in HRCT combing rule-based and support vector machine (SVM) classifiers is proposed for detecting different types of lung nodules, including both solid nodules and ground-glass opacity(GGO) nodules(part solid and nonsolid). This method consists of several steps: segmentation of pulmonary parenchyma and candidate pulmonary nodules, selection and extraction of features and classification of lung nodules, etc. This study demonstrates the superiority of the proposed method. Experimental results of recognition for lung nodules show desirable performances of the proposed method. The experimental results using the proposed method indicate the performances with sensitivity of 88.89% and specificity of 95.65%.

However, to further improve the detection performance, some improvements need to be further investigated as follows: (1)although some common features are selected and extracted in this paper, some more useful features, including 3D features, need to be further investigated and extracted; (2) in order to recognize lung nodules in noisy image more effectively, an adaptive smoothing method needs to be further

investigated; (3) From Table 2, it can be concluded that AUC is not the best criterion for grid search, but just a fairly good criterion. It is still a problem to solve that finding a best criterion for grid search. (4)in this paper, intelligent recognition of lung nodules are studied, but intelligent differentiating benign and malignant nodules is unconcerned with, which need to be further investigated in detail.

## 7. Acknowledgments

This work is supported by the National Engineering Research Center for Tissue Restoration & Reconstruction and the Guangdong Key Laboratory for Biomedical Engineering(South China University of Technology of China), Guangdong-Hong Kong Technology Cooperation Funding under Grant(No. 2010Z11), Cooperation project of Industry, Education and Academy, sponsored by Guangdong province government and Education Department of Chinese government(No.2009B090300057), Special funds to finance operating expenses for basic scientific research of Central Colleges(South China University of Tech., No.2011ZZ0021).

## References

1. J.D. Minna and J.H. Schiller, *Harrison's Principles of Internal Medicine* (McGraw Hall, Columbus, OH, 2008).
2. I. Guessous, J. Cornuz and F. Paccaud, Lung cancer screening: current situation and perspective, *Swiss Med Wkly.* 137(21/22) (2007)304–311.
3. X.J. Ye, X.Y. Lin and J. Dehmeshki, et al., Shape-based computer-aided detection of lung nodules in thoracic CT images, *IEEE Trans. Biomedical Engineering.* 56(7) (2009)1810–1820.
4. I. Sluimer, A. Schilham and M. Prokop, et al., Computer analysis of computed tomography scans of the lung: a survey, *IEEE Trans. Med. Imaging.* 25(4) (2006)385–405.

5. M. Gomathi and P.P. Thangaraj, Automated CAD for Lung Nodule Detection using CT Scans, in *2010 International Conference on Data Storage and Data Engineering*, (Bangalore, India, 2010), pp. 150–153.
6. T. Zrimec and S. Busayarat, A System for Computer Aided Detection of Diseases Patterns in High Resolution CT images of the Lungs, in *Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07)*, (Maribor, Slovenia, 2007), pp. 41–46.
7. D. S. Paik, C. F. Beaulieu and G. D. Rubin, et al., Surface normal overlap: a computer-aided detection algorithm with application to colonic polyps and lung nodules in helical CT, *IEEE Trans. Med. Imaging*. 23(6) (2004)661–675.
8. M. N. Gurcan, B. Sahiner and N. Petrick, et al., Lung nodule detection on thoracic computed tomography images: preliminary evaluation of a computer-aided diagnosis system, *Med. Phys.* 29(11) (2002)2552–2558.
9. S. G. Armato and W. F. Sensakovic, Automated lung segmentation for thoracic CT, *Acad. Radiol.* 11(9) (2004)1011–1021.
10. D.L. Pham, C.Y. Xu and J.L. Prince, A survey of current methods in medical image segmentation, *Annual Review of Biomedical Engineering*. (2)(2000)315–337.
11. S. Hu, E. A. Hoffman and J. M. Reinhardt, Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images, *IEEE Trans. Med. Imaging*. 20(6)(2001)490–498.
12. J. K. Leader, B. Zheng and R. M. Rogers, et al., Automated lung segmentation in X-ray computed tomography, *Acad. Radiol.* 10(11)(2003)1224–1236.
13. G. Q. Wei, L. Fan and J. Qian, Automated lung segmentation in X-ray computed tomography, in *Lecture Notes in Computer Science, Medical Image Computing and Computer-Assisted Intervention*, (Berlin, Germany 2002), vol.2488, pp. 746–752.
14. A.A. Enquobahrie, A. P. Reeves, and D. F. Yankelevitz, et al., Automated detection of pulmonary nodules from whole lung helical CT scans: performance comparison for isolated and attached nodules, in *Proc. SPIE*, (2004), pp. 791–800.
15. B. Zhao, M. S. Ginsberg, and R. A. Lefkowitz, et al., Application of the LDM algorithm to identify small lung nodules on low-dose MSCT scans, in *Proc. SPIE*, (2004), vol.5370, pp. 818–823.
16. K. Awai, K. Murao, and A. Ozawa, et al., Pulmonary nodules at chest CT: effect of computer-aided diagnosis on radiologists' detection performance, *Radiology*. 230(2)(2004)347–352.
17. Y. Lee, T. Hara, and H. Fujita, et al., Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique, *IEEE Transactions on Medical Imaging*. 20(7)(2001)595–604.
18. R. Wiemker, P. Rogalla, and A. Zwartkruis, et al., Computer aided lung nodule detection on high resolution CT data, *Proc. SPIE*, (2002) vol.4684, pp. 677–688.
19. S. Chang, H. Emoto, and D. N. Metaxas, et al., Pulmonary micronodule detection from 3-D chest CT, in *Medical Image Computing and Computer-Assisted Intervention* (2004) vol. 3217, pp. 821–828.
20. Q. Li and K. Doi, New selective enhancement filter and its application for significant improvement of nodule detection on computed tomography, in *Proc. SPIE*, (2004) vol. 5370, pp.1–9.
21. T. Chan and L. Vese, Active contours without edges, *IEEE Trans Image Process.* 10(2) (2001)266–277.
22. L. Vese, and T. Chan, A multiphase level set framework for image segmentation using the mumford and shah model, *Inter. Journal Computer Vision.* (50) (2002)271–293.
23. C.M. Li, C. Kao, and J. C. Gore, et al, Implicit active contours driven by local binary fitting energy, in *Computer Vision and Pattern Recognition, 2007. CVPR '07* (2007), pp.1–7.
24. C.M. Li, C. Y. Kao, and C.G. John, et al., Minimization of region-scalable fitting energy for image segmentation., *IEEE Trans. Image Processing.* 17(10) (2008)1940–1949.
25. Q. Li, F. Li, and K. Doi, Computerized detection of lung nodules in thin-section CT images by use of selective enhancement filters and an automated rule-based classifier. *Acad. Radiol.* (15) (2008)165–175.
26. C. C. McCulloch, R. A. Kaucic, and P. R. Mendonça, et al., Model-based detection of lung nodules in computed tomography exams. *Acad. Radiol.* 11(3) (2004)258–266.
27. M. S. Brown, M. F. McNitt-Gray, and J. G. Goldin, et al., Patient-specific models for lung nodule detection and surveillance in CT images. *IEEE Trans. Med. Imaging.* 20(12) (2001)1242–1250.
28. M. Tanino, H. Takizawa, and S. Yamamoto, et al., A detection method of ground glass opacities in chest x-ray CT images using automatic clustering techniques, in *Proc. SPIE*, (2003) vol. 5032, pp.1728–1737.
29. K. Awai, K. Murao, and A. Ozawa, et al., Pulmonary nodules at chest CT: effect of computer-aided diagnosis on radiologists detection performance. *Radiology*. 230(2) (2004)347–352.
30. D.Y. Lin, C.R. Yan, and W.T. Chen, Autonomous detection of pulmonary nodules on CT images with a neural network-based fuzzy system. *Comput. Med. Imaging Graph.*. (29) (2005)447–458.
31. Ei Baz, Ay man, and Gimel farb, et al., A new CAD system for early diagnosis of detected lung nodules, in *Image Processing, ICIP 2007. IEEE International Conference.*(San Antonio, TX, 2007), pp. 461–464.
32. Vapnik V, *The Nature of Statistical Learning Theory*. (Springer-Verlag Hall, New York, 1995).
33. Q. Zhao, and J.C. Principe, Support vector machines for SAR automatic target recognition. *IEEE Trans. Aerospace and Electronic System.* 37(2) (2001)643–654.
34. B. Scholkopf, A. Smola, and K.R. Muller, Nonlinear component analysis as kernel eigenvalue problem. *Neural Computation.* 10(5) (1998)1299–1319.



35. M. Kakar, and D. R. Olsen, Automatic segmentation and recognition of lungs and lesion from CT scan of thorax. *Comput. Med. Imaging Graph.* (33) (2009)72–82.
36. D. Wormanns, and D. Stefan, Characterization of small pulmonary nodules by CT. *European Radiology*. 14(1) (2004)1380–1391.
37. J. M. Goo, Computer-aided detection of lung nodule on chest CT: Issues to be solved before clinical use. *J. Radiol.* 6(2) (2005)62–63.
38. J. C. Platt, Sequential Minimal Optimization: a fast algorithm for training support vector machines. *Technical Report MSR-TR-98-14*. (1998).
39. R. E. Fan, P. H. Chen and C.J. Lin, Working set selection using second order information for training support vector machines. *Journal of Machine Learning Research*. (6) (2005)1889–1918.
40. E. Osuna, R. Freund and F. Girosi, Support vector machines: training and applications. *Massachusetts Institute of Technology, Cambridge, MA, USA*, (1997).
41. Y. S. Lee, Y. Ma and G. Jegadesh, Rolling-ball method and contour marching approach to identifying critical regions for complex surface machining. *Computers in Industry*.41 (2) (2000)163–180.
42. D. Barash, and T. Schlick, Multiplicative Operator Splittings in Nonlinear Diffusion: From Spatial Splitting to Multiple Timesteps. *Journal of Mathematical Imaging and Vision*. 19 (1) (2003)33–48.
43. K. Kanazawa, Y. Kawata, and N. Niki, Computer-aided diagnosis for pulmonary nodules based on helical CT images. *Compute Medical Image Graphcs*. 2 (2) (1998)157–167.
44. Y.C. Tang, Y. Q. Zhang, and V. Nitesh, et al., SVMs modeling for highly imbalanced classification. *Journal of Latex Class Files*. 1 (11) (2002)1–9.