

Research Article

Research on a Pulmonary Nodule Segmentation Method Combining Fast Self-Adaptive FCM and Classification

Hui Liu,^{1,2} Cai-Ming Zhang,^{1,2} Zhi-Yuan Su,^{1,2} Kai Wang,³ and Kai Deng⁴

¹*School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan 250014, China*

²*Digital Media Technology Key Lab of Shandong Province, Jinan 250014, China*

³*Lawrence Berkeley National Lab, University of California, Berkeley, CA 94720, USA*

⁴*Respiratory Department, Shandong Provincial Qianfoshan Hospital, Jinan 250014, China*

Correspondence should be addressed to Hui Liu; liuh_lh@126.com

Received 9 April 2014; Revised 1 November 2014; Accepted 21 December 2014

Academic Editor: William Crum

Copyright © 2015 Hui Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The key problem of computer-aided diagnosis (CAD) of lung cancer is to segment pathologically changed tissues fast and accurately. As pulmonary nodules are potential manifestation of lung cancer, we propose a fast and self-adaptive pulmonary nodules segmentation method based on a combination of FCM clustering and classification learning. The enhanced spatial function considers contributions to fuzzy membership from both the grayscale similarity between central pixels and single neighboring pixels and the spatial similarity between central pixels and neighborhood and improves effectively the convergence rate and self-adaptivity of the algorithm. Experimental results show that the proposed method can achieve more accurate segmentation of vascular adhesion, pleural adhesion, and ground glass opacity (GGO) pulmonary nodules than other typical algorithms.

1. Introduction

In recent years, morbidity and mortality rates of lung diseases in China are growing annually, resulting from environmental degradation induced by air pollution and increase of smoking and passive smoking. As indicated by data provided by National Cancer Institute, Ministry of Health, the mortality rate of lung cancer in China increased by 465% and its morbidity rate increased by 26.9% in the past 30 years, and furthermore, rather than liver cancer, lung cancer has been the prime cause of death due to malignant tumors, which takes up 22.7% of all such deaths (morbidity and mortality due to lung cancer increase drastically, <http://health.sohu.com/20121127/n358789664.shtml>). Therefore, research of computer-aided diagnosis on lung cancer has being increasingly important.

Currently, as far as the characterization of lung diseases is concerned, CT imaging is the best imageological means to diagnose lung diseases among various medical imaging types [1]. Since pulmonary nodules are the most commonly seen symptoms in the early stage of lung cancer, automatic

extraction and assisted diagnosis of pulmonary nodules by means of spiral CT play a significant role in the detection of early lung cancer [2], which reduces effectively the errors in interpreting images by eyeballs and the work load of image reading. More importantly, it improves the repeatability of diagnosis and consistency of image interpretation to some extent by providing accurate quantitative analysis for doctors by techniques such as image segmentation and quantification of characteristics [3].

In the process of research, a series of characteristics of pulmonary nodules such as size, form, density, and patterns of strengthening and growth need to be analyzed comprehensively in order to provide a basis for identification of benignity and malignancy of tumors [4]. For instance, pulmonary nodules have clear edges and the form of edges (smooth, lobulated, speculated protuberance, or burr) are key indicators of benignity and malignancy of tumors. So it cannot be more important to implement correct segmentations of pulmonary nodules.

In recent years, a lot of pulmonary nodule segmentations methods have been proposed, which can be categorized

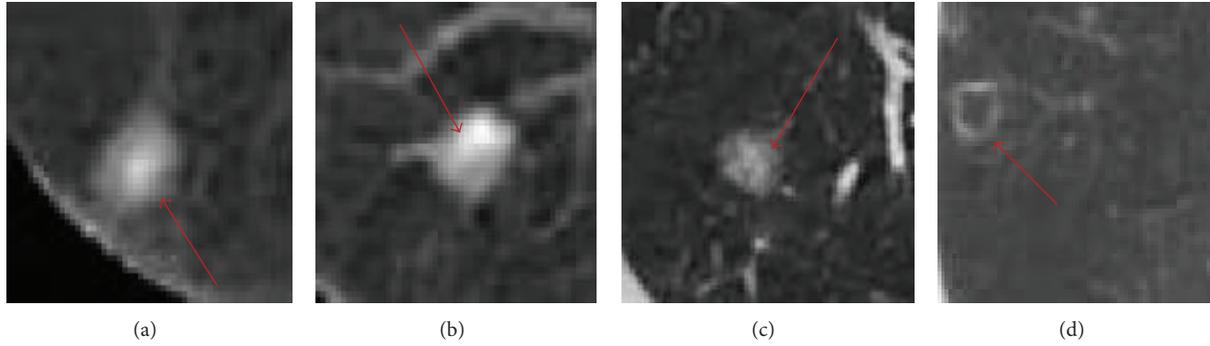


FIGURE 1: All kinds of pulmonary nodules: (a) vascular adhesion, (b) pleural adhesion, (c) ground glass opacity, and (d) cavitory nodule.

as threshold methods [5, 6], morphological methods [7], variable model methods [8], clustering methods [9–11], and image segmentation methods [12, 13]. Nevertheless, many contemporary methods need human interventions and most of them perform well only for isolated solid pulmonary nodules. However, pulmonary nodules usually occur in forms of vascular adhesion, pleural adhesion, ground glass opacity (GGO), and cavitory pulmonary nodules, as demonstrated in Figure 1. These types of nodules have higher probability of malignancy than isolated ones. Obviously, because of attachment of these nodules to surrounding tissues, fuzzy edges, and similar gray scales, fast, correct, and automatic segmentation of these nodules is still a difficult problem.

Fuzzy C -means (FCM) algorithm [14] is a popular algorithm and has been widely used in image segmentation; it applies fuzziness to the membership judgment of pixels, which is consistent with human perception and convenient for realization. However, conventional FCM algorithm does not consider any spatial information and neighborhood correlation, which makes it perform poor in low contrast, inhomogeneity, and noisy images.

In order to overcome these problems, many researchers have introduced spatial information through modifying the objective function or altering the distance form measurement between pixels and cluster centers. Fergus et al. proposed a modified FCM algorithm (SFCMpq), which considers the fact that the adjacent pixels will have similar characteristic value and the probability that they belong to the same class is very big. Chuang et al. [15] proposed a weighted FCM integrating the spatial neighborhood information. Ahmed et al. [16] proposed FCM_S algorithm where the objective function is modified by introducing a neighborhood term, which is used to compensate the intensity inhomogeneity and allow the labeling of a pixel to be influenced by the labels of its immediate neighborhood. However, the execution efficiency of FCM_S is very low [17], since it needs to compute a series of neighborhood labelings during each iteration. Two improved variants, FCM_S1 and FCM_S2, of FCM_S were proposed with an intent to simplify the computation of parameters and then Chen and Zhang [18] extend them to the corresponding kernelized versions, KFCM_S1 and KFCM_S2, by the kernel function substitution. An enhanced FCM (EnFCM) algorithm [19] was proposed to speed up the

clustering process on the basis of the gray level histogram instead of pixels, and it uses a linearly weighted sum image formed from both original image and each pixel's local neighborhood average gray level. Hence, the computational time of EnFCM algorithm is reduced greatly since the number of gray levels in an image is much smaller than that of pixels. Fast generalized FCM (FGFCM) algorithm [20] introduces the spatial information combining the intensity of the local pixel neighborhood and the number of gray levels in an image. The quality of segmentation result is well enhanced, and the computational time of FGFCM algorithm is small similarly. Krinidis and Chatzis proposed a novel robust fuzzy local information C -means clustering (FLICM) algorithm [21] by incorporating the local spatial and gray level information, which is free of the empirically adjusted parameters and enhances the clustering performance.

Although the above-modified FCM algorithms have improved the performance of conventional FCM and segmented different types of images to some extent, they still have the following disadvantages: (1) they require a large number of iterations to achieve the convergence; (2) they have the worst performance in segmenting CT images of lungs with complicated features; for example, the blood vessels have similar grayscales to pulmonary nodules, GGO pulmonary nodules have low contrast to backgrounds, and so forth.

Discussing these issues, this paper presents a pulmonary nodule segmentation method combining fast self-adaptive FCM and classification. The improved spatial function will calculate the cluster centers adaptive by both grayscale similarity and spatial similarity of pixels neighborhood space and update the fuzzy membership degree of center pixel during each iteration, obtaining the cluster centers at a faster rate. Afterward, we can construct the classifier based on the clustering results according to the labeled information.

2. Modified Fast Self-Adaptive FCM Segmentation

In addition to lung tissues, a typical CT image of chest section includes other organs, skeletons, and trachea. To increase the accuracy of CAD detection to lesion tissues and image processing speed, the primary mission of this paper is to remove these irrelevant items in images and extract images

of pulmonary parenchyma by means of global binarization of threshold, detecting boundaries of trunks, removing background noises, and inpainting lung areas. Thereafter, nodules in lungs as well as blood vessels and bronchi which have similar characteristics with pulmonary nodules can be analyzed. CT images of pulmonary nodules are generally bright in center and dark in peripheral area so that points mapped into the joint characteristic space of brightness and greyscales are dense in center and scattered in peripheral. Considering the diversity of types of pulmonary nodule density distribution, weak edge distribution of peripheral tissues, and robustness of algorithm to noise, the following modifications of traditional FCM algorithm are proposed in this paper: the new spatial function considers the contributions to fuzzy membership from both the grayscale similarity between central pixels and single neighboring pixels and the spatial similarity between central pixels and neighborhood. The modified algorithm improves effectively the convergence rate and adaptivity to spatial characteristics of pulmonary nodule edges.

2.1. The Traditional FCM Algorithm. As a clustering algorithm based on division, the traditional FCM algorithm divides N pixels into C fuzzy groups, searches the clustering center of each group, and updates the clustering centers through updating fuzzy membership of the pixels relative to each clustering. The update is implemented by means of minimizing the objective function, which is defined as

$$E = \sum_{i=1}^C \sum_{j=1}^N \mu_m^{ij} \|x_j - v_i\|^2, \quad (1)$$

where C is the amount of clusters, N is the amount of pixels in the image, μ_m^{ij} is the fuzzy membership of the j th pixel with respect to the i th pixel, m is the weight exponent acting on the fuzzy membership, and v_i is the center of the i th clustering. The constraint for the objective function minimization is

$$\mu_m^{ij} \in [0, 1], \quad \sum_{i=1}^C \mu_{ij} = 1. \quad (2)$$

Differentiating equation (1) with respect to μ_m^{ij} and v_i obeying constraint (2) yields yields a necessary condition of minimizing (1):

$$\mu_m^{ij} = \frac{1}{\sum_{i=1}^C [\|x_j - v_i\| / \|x_j - v_k\|]^{2/(m-1)}}, \quad (3)$$

$$v_i = \frac{\sum_{j=1}^n \mu_m^{ij} x_j}{\sum_{j=1}^n \mu_m^{ij}}. \quad (4)$$

This algorithm minimizes the objective function by assigning higher fuzzy membership to pixels whose greyscales are close to those of specific clustering centers and assigning lower fuzzy membership to pixels whose greyscales are distinct from those of specific clustering centers.

Nonetheless, the traditional FCM algorithm calculates the fuzzy membership of each pixel based on the distance of

the pixel and clustering centers, resulting in the fact that it is not suitable to pulmonary nodule images in which characteristics of forms and distributions are complex and diverse because it is difficult to obtain smooth segmentation edges and closed segmentation intervals. In addition, the executive efficiency of the FCM algorithm is low since it needs multiple iterations to converge. Existing modified FCM algorithms, such as FCM_S, FCM_S1, FCM_S2, FGFCM, and EnFCM, take into account information of pixels in neighborhoods of central pixels and introduce a parameter α or λ_g that controls the intensity of neighborhood information. This parameter is important because its selection influences directly the performance of algorithm. However, this parameter cannot vary self-adaptively; that is, its value maintains unchanged during the whole process of segmentation for an image. Obviously, the selection of α does not reflect the characteristics of the neighborhood of the pixels.

We acclaim that the value of α should be chosen based on (1) similarity of local greyscales; that is, α should depend on the difference between the greyscales of central pixels and neighboring pixels in order to suppress noise and speed convergence; (2) local spatial similarity; that is, α also depends on the size of neighborhoods and the difference between greyscales of neighborhood average and central pixels so that the adaptivity of algorithm is enhanced.

2.2. 2D Vector Representation of the Spatial Relationship between a Pixel and Its Neighborhood. For an extracted pulmonary parenchyma image, the amount of pixels is denoted as N and grayscale lies in the interval $[0, L - 1]$, the neighborhood of a pixel x_j , $j \in [0, N]$, is N_r , and x_r is a pixel in the neighborhood. Denote the grayscale at the central pixel x_j as $f(x_j) \in [x, L - 1]$ and the average value in a neighborhood centered at this pixel with the size $(2r + 1) \times (2r + 1)$ as $g(x_j)$, where the parameter r is determined based on image resolution and noise. The 2-dimensional vector $[f(x_j), g(x_j)]$ represents the grayscale at x_j and that of its neighborhood; it will have M^* different values in the whole lung parenchyma. The algorithm segments the image by sampling the above 2D vector instead of the traditional grayscale.

To analyze the statistical rules of the correspondence between 2D vectors and various tissues in pulmonary parenchyma, a 2D grayscale matrix is defined and constructed, in which the abscissa represents the greyscales at pixels and the ordinate represents the average in their neighborhoods. The frequency p_j of the vector value $[f(x_j), g(x_j)]$ is

$$p_j = \frac{n_j}{N}, \quad \sum_{j=1}^{M^*} p_j = 1, \quad (5)$$

where n_j means the number of times that the vector value $[f(x_j), g(x_j)]$ occurs.

Figure 2(b) shows the distribution of greyscales and neighborhood averages of pixels corresponding to pulmonary parenchyma shown in Figure 2(a), for which r is chosen as 2. By statistical analysis from the 2D vector distributions at all pixels, since pulmonary parenchyma is the

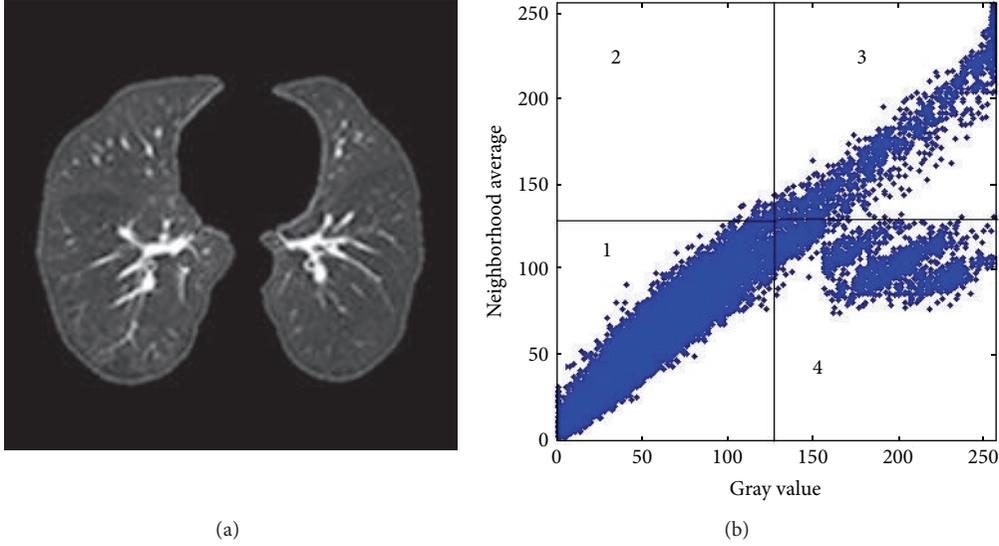


FIGURE 2: Greyscale and neighborhood average: (a) a pulmonary parenchyma image and (b) 2D vector distribution; regions 1 and 3 correspond to the smooth area, region 4 corresponds to the edges of nodules, tracheas, great blood vessels, and so forth.

background in images, the greyscale values at corresponding pixels are close to their neighborhood averages, which is indicated by points around the diagonal in region 1 in Figure 2(b); greyscale values at pixels inside tracheas, blood vessels, and pulmonary nodules are also close to their neighborhood averages, indicated by points around the diagonal in region 3 in Figure 2(b); amount of pixels at edges of nodules, tracheas, and great blood vessels is relatively large and greyscales at these areas are significantly different from neighborhood averages, shown by points densely distributed in region 4. In addition, discrete points showing low frequency of occurrence and huge difference between local greyscale values and neighborhood averages correspond to pixels at tiny blood vessel edges and lung wall edges.

From above, for a pixel, the size of its neighborhood, the difference between local greyscale and neighborhood average, and its frequency p can provide local pattern features of the neighborhood around the central pixel for the algorithm updating fuzzy membership.

2.3. FCM Algorithm Based on Weighted Spatial Information and Greyscale Information. To improve the traditional FCM algorithm, this paper introduces a new spatial function which includes the spatial features and greyscale information of the neighborhood around the central pixel. The spatial function is defined as

$$h_{ij} = \sum_{x_r \in N_r} \mu_{ir} + \omega_r, \quad (6)$$

where h_{ij} represents the probability that the pixel determined from spatial information of the neighborhood belongs the i th kind and μ_{ir} is the membership that the neighboring pixels x_r belong to the i th kind, which is calculated by (3). Obviously, h_{ij} is maximum if all pixels in the neighborhood of the pixel x_j belong to the i th kind. Otherwise, h_{ij} is minimum if all

pixels in the neighborhood of the pixel do not belong to the i th kind.

A polynomial ω_r represents the global scale factor determined by the difference between greyscales at the pixel x_j and neighboring pixels x_r ; that is, the spatial function h_{ij} depends on the fuzzy membership of neighboring pixels x_r , and more importantly contributions to the spatial function h_{ij} from different neighboring pixels x_r around x_j should be distinct. The less the greyscales at x_j and x_r are different, the closer their relation is and the greater the corresponding weight coefficient should be. Otherwise, the weight coefficient should be smaller. Thus, ω_r is defined as

$$\omega_r = \frac{1}{\exp(|x_j - x_r|^2) \cdot \lambda_j}, \quad (7)$$

where λ_j is equivalent to the parameter α in other classical algorithms. But λ_j is able to adjust the difference between greyscales at central pixels and their neighborhoods according to the degree of importance of the 2D vector in local pattern space. Consider

$$\lambda_j = p_j \cdot \sqrt{\sum_{x_r \in N_r} \frac{(f(x_r) - g(x_j))^2}{(2r + 1)^2}} \quad (8)$$

which shows that λ_j is variable and enhances the self-adaptivity of pixels to their neighborhoods.

It can be seen that formula (7) combines the greyscales at central pixels and pixels in their neighborhoods and adjust dynamically the brightness around central pixels depending on the sizes of neighborhoods and mean square deviation in neighborhoods. This practice protects the pixels in smooth areas, magnifies the differences in edges, and eliminates noise and edges of tiny suspect objects in the fuzzy membership matrix.

The spatial function h_{ij} is plugged into the following formula in order to update the fuzzy membership of the pixel x_j :

$$\mu'_{ij} = \frac{\mu_{ij}^p \cdot h_{ij}^q}{\sum_{k=1}^C \mu_{ij}^p \cdot h_{ij}^q}, \quad (9)$$

where parameters p and q control the relative significance of the two functions.

Based on the above descriptions, the self-adaptive FCM algorithm for training is as follows:

Algorithm 1 (fast self-adaptive FCM). Consider the following.

Input. Extract CT image of pulmonary parenchyma.

Output. Segment pulmonary nodules.

Initializations. Choose the amount of clustering centers $C = 2$, the maximum number of iterations $L_{\max} = 100$, fuzzy weight exponent $m = 2$, size parameter of neighborhoods $r = 2$, parameters that control the relative significance of the two functions $p = 1$ and $q = 2$, and termination threshold $\varepsilon = 0.001$; set counter for iteration l_0 ; initialize randomly a clustering prototype matrix $V^{(0)}$ and normalize this matrix so that it satisfies constraint (2); finally, obtain clustering centers $v_i^{(0)}$ using formulas (3) and (4).

Step 1. Calculate and plot the 2D histogram for the pulmonary parenchyma image and obtain frequencies of occurrence p_j of all 2D vectors.

Step 2. Calculate fuzzy membership $\mu_{ij}^{(l)}$ of pixel x_j and assemble the partition matrix $U^{(l)}$:

$$\forall i, j, \exists |x_j - v_i|^{(l)} > 0, \text{ then} \quad (10)$$

$$\mu_{ij}^{(l)} = \left\{ \sum_{k=1}^C \left[\frac{|x_j - v_i|^{(l)}}{|x_j - v_k|^{(l)}} \right]^{2/(m-1)} \right\}^{-1}.$$

If $\exists i, r$, such that $|x_j - v_i|^{(l)} = 0$, then $\mu_{ir}^{(l)} = 1$ and $(\mu_{ir}^{(l)} = 0$ for $i \neq k$).

Step 3. Update fuzzy memberships according to formulae (7)–(9) and obtain new fuzzy memberships $\mu_{ij}^{(l)}$.

Step 4. Update clustering prototype modes $V^{(l)}$; that is, update clustering centers $v_i^{(l+1)}$ using new memberships $\mu_{ij}^{(l)}$:

$$v_i^{(l+1)} = \frac{\sum_{j=1}^N \mu_{ij}^{(l)m} (x_j + (1/N_R) \sum_{x_r \in N_R x_r})}{\sum_{j=1}^N \mu_{ij}^{(l)m}}, \quad (11)$$

$$i = 1, 2, \dots, C.$$

Step 5. If $\|V^{(l+1)} - V^{(l)}\| < \varepsilon$ or the maximum count of iteration L_{\max} is reached, enter Step 6; otherwise, return to Step 3.

Step 6. Deblur using the method of maximum membership function. The type of clustering that the pixel x_j belongs to is represented by c_j :

$$c_j = \arg \left\{ \max_i (\mu_{ij}) \right\} \quad \forall i, \forall k. \quad (12)$$

In the algorithm, the fuzzy weight exponent m is a parameter introduced to generalize fuzzy clustering objective functions by Bezdek, which plays a role in suppressing noise and smoothing membership functions. The empirical range of m given by Bezdek is $1.1 \leq m \leq 5$. Later, the most suitable physical interpretation was obtained for $m = 2$ [22]. Thus m is chosen as 2 in this paper.

Now, let us analyze the number of iterations and running time of the algorithm for image processing. Since the iterations loop over each pixel, the time complexity of the algorithm depends closely on amount of pixels and number of iterations, that is, $O(NcL)$, where N is the amount of pixels, c is the number of clusters, and L is the number of iterations. The operational speed of the algorithm in this paper is significantly improved because the number of iterations of this algorithm is much less than traditional FCM and FCM_S algorithms. Besides, EnFCM and FGFCM algorithms transform segmentation based on pixels into segmentation based on greyscale (0–255) levels. Due to the fact that the number of greyscale levels is much less than amount of pixels, the speed of algorithm becomes much faster.

Figure 3 demonstrates the results of fuzzy memberships of pixels at pulmonary parenchyma, edges, and pulmonary nodules given by the above algorithm applied to an image of blood vessel adhesion pulmonary nodules. The yellow rectangular area in Figure 3(a) is the area to be analyzed. Figure 3(b) shows the distribution of greyscales at pixels in the rectangular area in the form of level sets. Figure 3(c) indicates the results of fuzzy memberships at the pixels in the interesting area relative to the foreground clustering c_F calculated by using the new spatial function. To make the demonstration of results more visually sense-making, the values shown in the figure are the results that actual values are multiplied by 100 and rounded. Black arrows point to the evolutionary directions of fuzzy memberships. Pixels in shadow areas represent segmented boundaries based on the membership threshold.

Let us elaborate the effectiveness of the new spatial function by taking the pulmonary parenchyma shown in Figure 4, for instance. The positions that arrows point to can be regarded as noise. The frequency p that the corresponding 2D vector occurs is extremely small and its greyscale differs significantly from most neighboring pixels. Thereafter, in the initial stage, μ_{ij} is relatively small while μ_{ir} is large, which lead to the result that h_{ij} given by the algorithm through the spatial function (6) is large and the new fuzzy memberships μ'_{ij} by substituting h_{ij} into formula (9) is also relatively large. After each iteration, the probability that this pixel enters the i th clustering increases until the termination of the algorithm. Thus, this pixel will be finally categorized into the i th clustering where its neighboring pixels belong to, although its greyscale is distinct from the center of the i th clustering v_i . It is

right opposite to the case of pleural adhesion nodules shown in Figure 4(b), in which the area of pulmonary nodules extends until being connected to pleura that shows higher greyscale levels. Though greyscale levels at the connections are low, they transit gradually so that the difference of greyscales at central pixels and neighborhoods is relatively small. The factor λ_j in formula (8) adjusts self-adaptively the brightness around pixels in the connection area by increasing the weight of μ_{ij} , which also increases the spatial function h_{ij} . After one iteration, fuzzy membership changes significantly.

- (1) The central pixel is a point of noise and pixels in its neighborhood are homogeneous in greyscale. Let us take a point of tiny blood vessel noise shown in Figure 4(a) as an example, in which the size parameter of neighborhood $r = 1$. Greyscales and initial fuzzy memberships of the pixels in this area are listed in Table 1(a). Tables 2(a), 2(b), and 2(c) demonstrate the memberships μ'_{ij} and clustering centers v_i of the pixels in this area after 1, 4, and 7 iterations. In Table 2, the upper and lower parts show results given by formula (1) and Step 3 in Algorithm 1, respectively. By comparing the memberships and clustering centers before and after updating, it can be seen that clustering centers after each iteration are rarely influenced by central pixels, speed of convergence is improved significantly, and fuzzy memberships of pixels in neighborhoods tend to be consistent. Thus the noise of tiny blood vessels is correctly categorized into the background clustering c_B .
- (2) The central pixel is not a point of noise but is homogeneous in greyscale with pixels in its neighborhood. Table 3 demonstrates the memberships and clustering centers of the pixels in the neighborhood of the example pixel shown in Table 1(b) after 1, 4, and 7 iterations. It can be seen that since the central pixel of the neighborhood is not noise, only the greyscales at diagonal pixels are relatively dark and the greyscales in the neighborhood show smooth transitions. Through comparing the memberships μ'_{ij} and clustering centers v_i before and after updating as listed in Tables 3(a), 3(b), and 3(c), it is observed that after one iteration in the algorithm proposed by this paper, memberships μ'_{ij} at pixels in the neighborhood are lower than those computed by Step 2 of Algorithm 1 and membership values tend to be consistent after several iterations. The central pixel is correctly categorized into the foreground clustering c_F .

The above two examples provide theoretical explanations about the robustness of the algorithm. From the point of view of localized methods, formula (6) fully combines the local spatial information and greyscales of the image, which enhances the robustness of the algorithm to noise or outliers and the self-adaptivity to homogenous neighborhoods. In the meantime, it can change the fuzzy memberships of central pixels significantly and improve the executive efficiency of the algorithm.

TABLE 1: Greyscales and initial memberships in the neighborhood of pixel x_j in Figure 6.

(a) Central pixel x_j is the noise of tiny blood vessel in Figure 7(a)

Greyscale			Initial membership		
78	53	53	0.22	0.11	0.39
60	142	52	0.82	0.13	0.25
60	59	56	0.42	0.54	0.25

(b) Central pixel x_j is connection of nodule and lung wall in Figure 7(b)

Greyscale			Initial membership		
132	142	162	0.27	0.89	0.78
140	135	168	0.55	0.18	0.44
153	156	138	0.54	0.84	0.42

TABLE 2: Values of memberships and clustering centers (v_1 and v_2) in the neighborhood shown in Figures 4(a) and 4(b) after 1, 4, and 7 iterations.

(a) After 1 iteration

0.4753	0.4727	0.4712
0.4717	0.4723	0.4753
0.4721	0.4736	0.4722
$v_1 = 80.54$	$v_2 = 37.46$	
0.5031	0.5122	0.5124
0.5112	0.53031	0.5042
0.5121	0.5063	0.5066
$v_1 = 83.55$	$v_2 = 42.68$	

(b) After 4 iterations

0.7564	0.7562	0.7601
0.7537	0.7523	0.7564
0.8120	0.7611	0.7589
$v_1 = 52.44$	$v_2 = 63.27$	
0.7781	0.7881	0.7689
0.7833	0.7682	0.7596
0.7681	0.8195	0.7764
$v_1 = 36.88$	$v_2 = 81.67$	

(c) After 7 iterations

0.9687	0.9311	0.9714
0.9658	0.9602	0.9596
0.9709	0.9673	0.9598
$v_1 = 20.74$	$v_2 = 98.65$	
0.9868	0.9727	0.9943
0.9907	0.9739	0.9763
0.9960	0.9902	0.9856
$v_1 = 7.29$	$v_2 = 118.69$	

3. Experimental Results and Analysis

This paper uses experimental data from whole lung image series provided by the Lung Image Database Consortium (LIDC (USA Lung Image Database Consortium (LIDC),

TABLE 3: Values of memberships and clustering centers (v_1 and v_2) in the neighborhood shown in Figure 7(b) after 1, 4, and 7 iterations.

(a) After 1 iteration		
0.6007	0.5507	0.5677
0.5236	0.5712	0.5178
0.5807	0.5008	0.5438
$v_1 = 82.25$	$v_2 = 30.58$	
0.4922	0.4993	0.4892
0.5934	0.4993	0.5372
0.4993	0.4898	0.5192
$v_1 = 71.45$	$v_2 = 44.86$	
(b) After 4 iterations		
0.3446	0.3247	0.3251
0.4024	0.3512	0.3415
0.3515	0.3861	0.3488
$v_1 = 46.28$	$v_2 = 76.55$	
0.2452	0.3412	0.2944
0.3544	0.2586	0.3589
0.2881	0.3245	0.2292
$v_1 = 34.62$	$v_2 = 82.54$	
(c) After 7 iterations		
0.0278	0.0852	0.0353
0.0267	0.0497	0.0318
0.0795	0.0278	0.0519
$v_1 = 20.48$	$v_2 = 100.56$	
0.0105	0.0084	0.0068
0.0097	0.0092	0.0087
0.0082	0.0081	0.0051
$v_1 = 5.33$	$v_2 = 118.36$	

<https://public.cancerimagingarchive.net/ncia/login.jsf>) [23], USA, and the lung CT image database of Qianfoshan Hospital, Shandong Province. Table 4 describes the two sets of experimental data, 2D CT images selected for experiments, and information about pulmonary nodules. The experimental images contain labels of pulmonary nodules provided by clinical experts.

To verify the effectiveness of the newly proposed algorithm, the new algorithm is compared with 5 existing typical algorithms (FCM, FCM.S, EnFCM, FGFCM, and FLICM) about effect of segmentation, operational efficiency, and error rate. In the whole experiment, parameters involved in the above 5 algorithms are as follows: $m = 2$, $\alpha = 0.85$, $\varepsilon = 0.001$, $\lambda_g = 6$, and $N_R = 9$; that is, local neighborhoods are 3×3 in size.

3.1. Effect of Pulmonary Nodule Segmentations. Segmentations to isolated, blood vessel adhesion, pleural adhesion, and ground glass opacity pulmonary nodules are implemented and compared in order to verify and compare the performance of the above algorithms on removing noise and treating tiny tissues, objects with fuzzy edges, and objects with similar greyscales.

TABLE 4: Databases used in the experiment.

	LIDC	Qianfoshan Hospital, Shandong Province
Average number of sections in each set	100	150
Number of 2D CT images for experiment	48	67
Image resolution	512×512	512×512
Thickness of sections	2~3 mm	1~3 mm
Diameter of pulmonary nodules	8~30 mm	7~30 mm
Number of pulmonary nodules		
Isolated	64	98
Blood vessel adhesion	41	56
Pleural adhesion	19	39
ground glass opacity	25	21

Figure 5 demonstrates the image of an isolated pulmonary nodule, the image of pulmonary parenchyma with salt and pepper noise, and images of locally zoomed-in segmentation results obtained by the above six algorithms and by manual segmenting. Through analysis of the segmentation results shown in Figures 5(c)–5(h), it can be found that the performances of FCM and FGFCM are influenced by noise so severely that they cannot separate objectives from background. A large amount of noise still exists in the result of background segmentation and even tiny blood vessels are segmented as objectives. FCM.S and EnFCM perform better than above algorithms, although the segmentation effects of objectives and tiny blood vessels are much less than ideal. The algorithm proposed by this paper can remove the influence of noise satisfactorily and successfully separate objectives from background.

Figure 6(a) shows the CT image of a blood vessel adhesion pulmonary nodule. The segmentation results of this nodule by the above six algorithms are demonstrated by Figures 6(b)–6(g). It is observed that, among all these algorithms, FCM yields a result that differs the most from that given by manual segmentation because it only extracts the brightest region in the center of the nodule and loses all information about nodule boundaries. Segmentations provided by FCM.S, EnFCM, FGFCM, and FLICM show errors of boundary leakage at the connections of nodules and blood vessels, and all of these four algorithms make more or less wrong segmentations to left lung walls connected to blood vessels. The method proposed by this paper overcomes these problems and performs well in segmenting blood vessel adhesion pulmonary nodules.

Figure 7(a) shows the CT image of a pleura adhesion pulmonary nodule. The segmentation results of this nodule by the above six algorithms are demonstrated for comparison by Figures 7(b)–7(g). The shape of pulmonary nodule edges segmented by our method are the closest to those provided by experts' manual segmentations. The connections between nodules and pleura are especially delicate, indicated by the

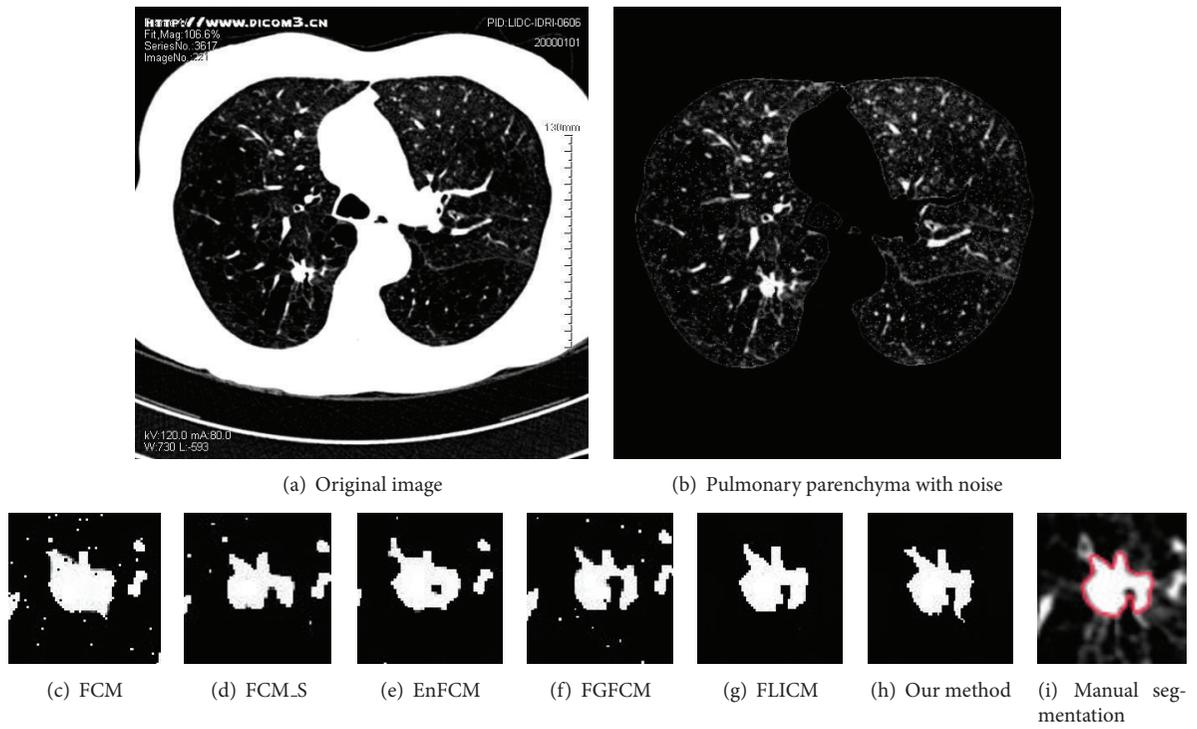


FIGURE 5: Segmentation results of isolated pulmonary nodules with salt and pepper noise.

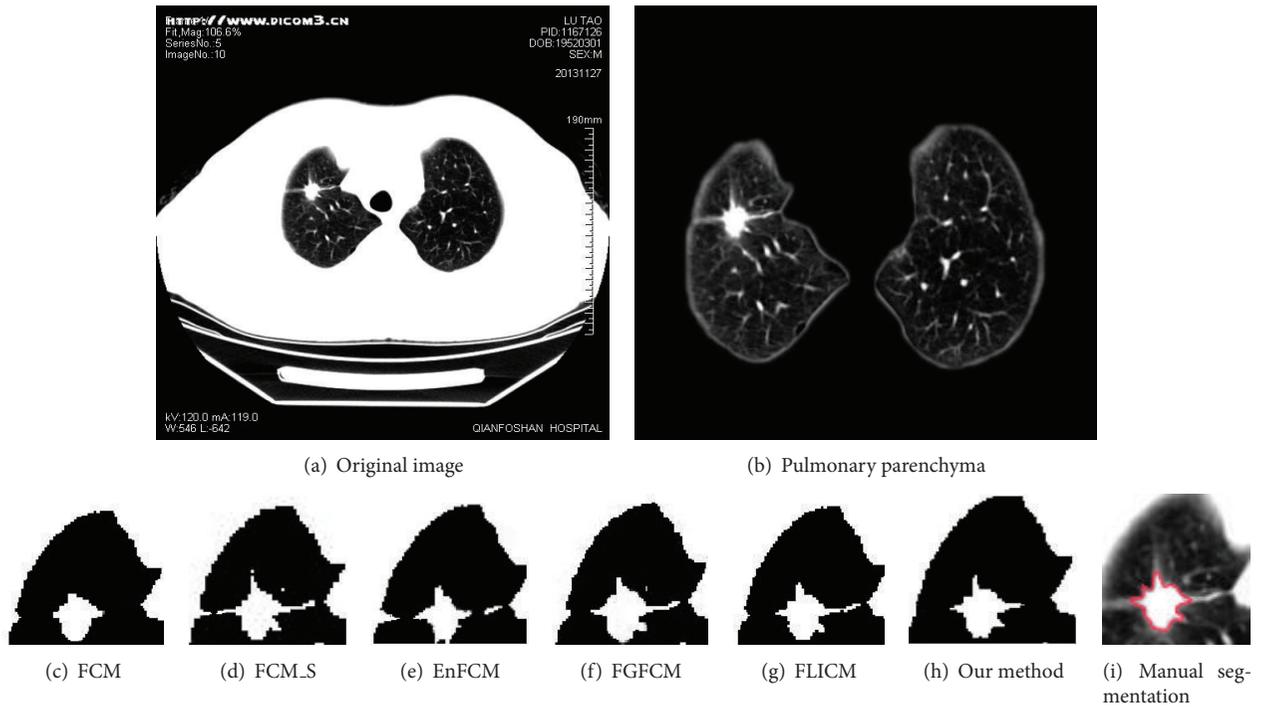


FIGURE 6: Segmentation results of blood vessel adhesion pulmonary nodules.

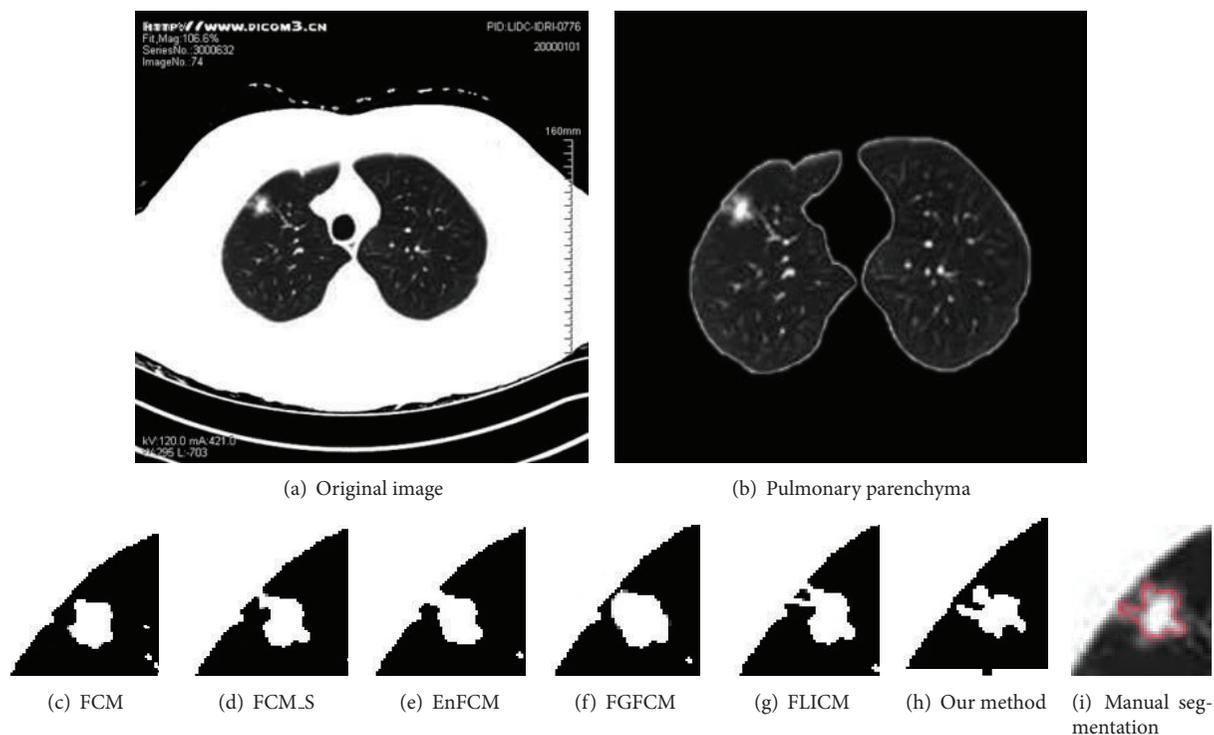


FIGURE 7: Segmentation results of pleura adhesion pulmonary nodules.

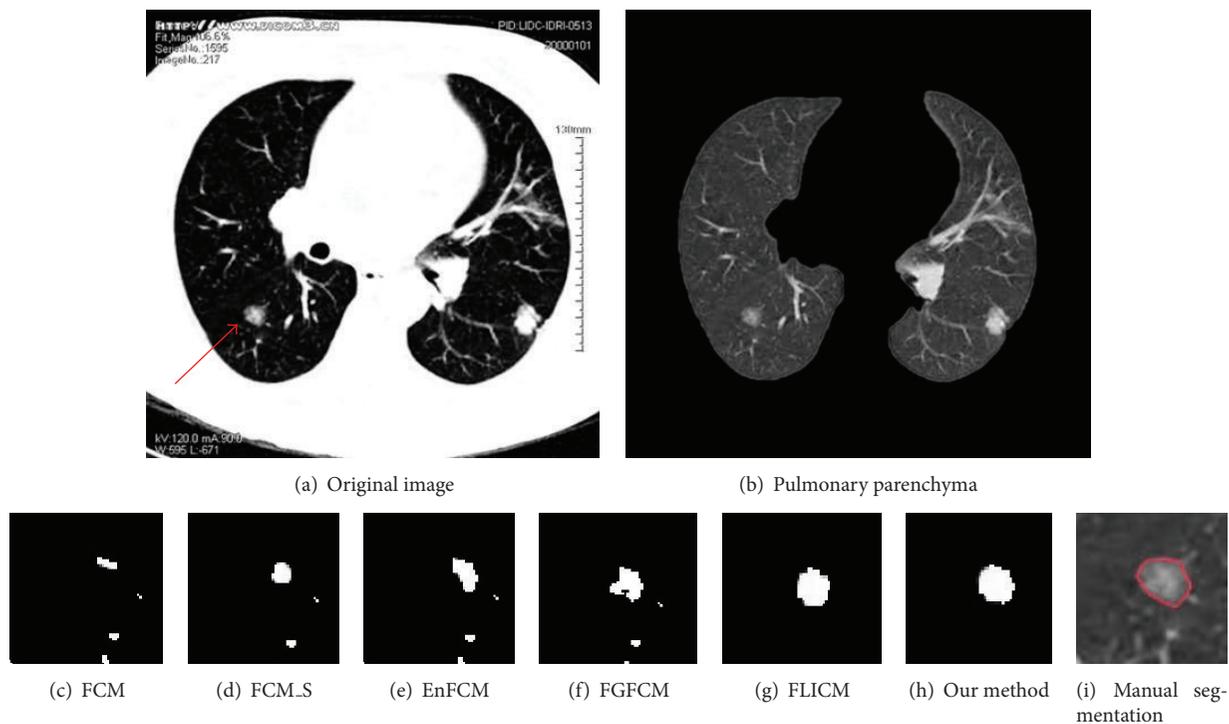


FIGURE 8: Segmentation results of ground glass opacity pulmonary nodules.

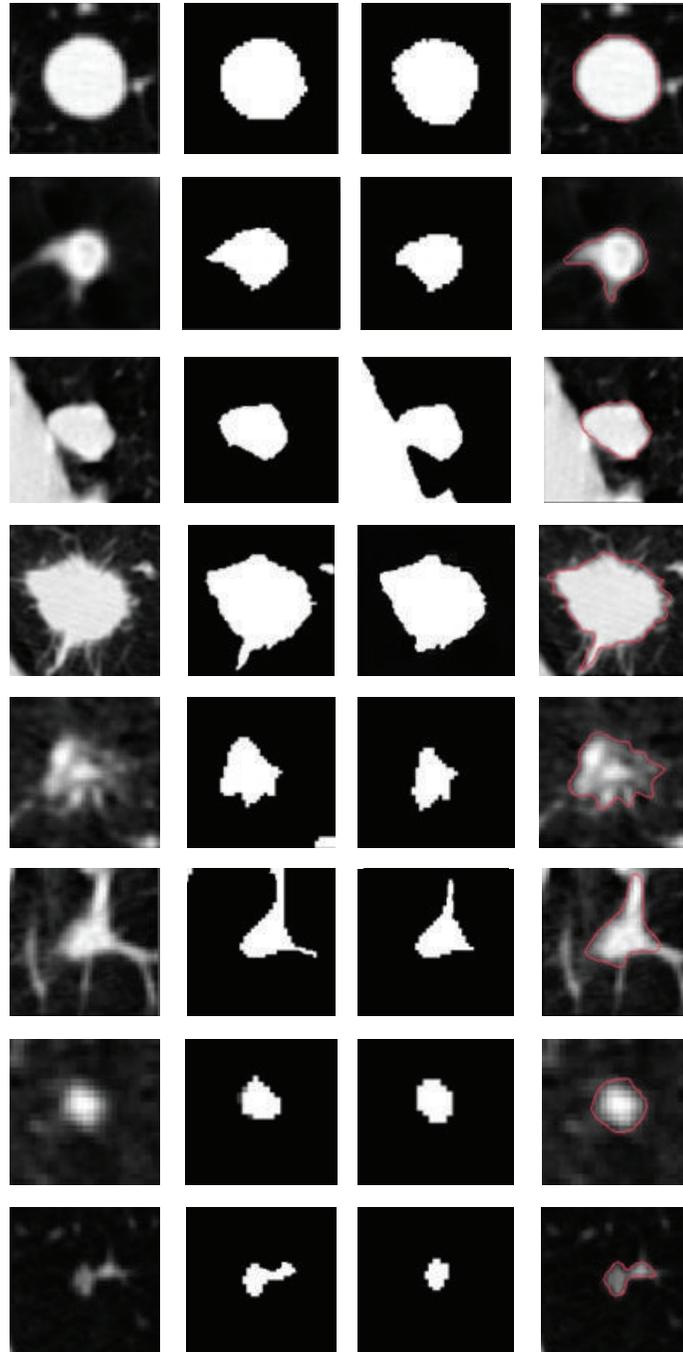


FIGURE 9: Examples of pulmonary nodule segmentations by our algorithm compared with manual segmentations. For each case, four images are shown. They are, from left to right, the original CT slice, segmentation result with $r = 1$ and $r = 2$, respectively, and manual segmentation result.

fact the extended area of nodules and background areas of pulmonary parenchyma are distinguished.

Figure 8(a) shows the CT image of a ground glass opacity pulmonary nodule, which is indicated by a red arrow. The segmentation results of this nodule by the above six algorithms are demonstrated for comparison by Figures 8(b)–8(g). It can be seen that FCM hardly segments pulmonary nodules. FCM_S, EnFCM, and FGFCM perform less than

satisfactorily on segmentations of nodules and tiny blood vessels. Segmentations by FLICM are good and our method gives segmentations that are closer to edges expected by experts.

Generally, blood vessels have similar grayscale to pulmonary nodules; GGO pulmonary nodules have low contrast to backgrounds; grayscales at connections between pleura and pulmonary nodules transit smoothly, but the contrast is also very low, compared to the grayscales of backgrounds.

TABLE 5: Comparison of executive efficiencies of the six algorithms.

Algorithm	Number of iterations			
	Figure 5 (noise added)	Figure 6	Figure 7	Figure 8
FCM	69	50	35	68
FCM.S	36	14	22	63
EnFCM	11	28	13	56
FGFCM	11	27	14	47
FLICM	18	33	13	15
Algorithm in this paper	6	8	5	9
	Running time (s)			
FCM	12.8605	12.4230	9.8213	17.2120
FCM.S	10.9752	9.1020	8.3464	14.3180
EnFCM	7.2450	8.3830	8.1840	12.3040
FGFCM	9.2880	7.4350	6.1859	9.3020
FLICM	6.2450	9.3020	6.1826	7.7290
Algorithm in this paper	5.3614	4.6481	4.4618	5.7344

The traditional FCM algorithm implements segmentations of images by clustering similar data points in the characteristic space and does not include interactions between neighboring pixels, which yields the worst performance in segmenting CT images of lungs with complicated features. Improvements by introducing neighborhood information and methods to measure distances and correcting memberships are made by all other four typical algorithms, which remove the influence of image noise. However, more or less defects exist in considering spatial similarities and grayscale similarities of pixels in neighborhoods so that these algorithms show problems of falling into local extrema and being sensitive to initial values.

To illustrate more results of segmentations of pulmonary nodules, Figure 9 demonstrates several typical pulmonary nodules, including isolated, adhesion, and GGO nodules. Column 1 shows locally zoomed-in images of original pulmonary parenchyma; column 2 shows segmentation results given by using the neighborhood size parameter $r = 1$, that is, using 3×3 neighborhoods; column 3 shows segmentation results given by using the neighborhood size parameter $r = 2$, that is, using 5×5 neighborhoods; column 4 shows manual segmentation results provided by experts for reference.

Analysis of experimental results demonstrates the following: (1) to some extent, the segmentation obtained by our method is close to the manual references, especially at nodule edges where burrs are obvious. (2) Images in the same group can be referred to each other and those in different groups cannot be compared because different CT images are distinct in doses, resolutions, and quality, especially grayscale, and in the process of implementation the algorithm adjusts the parameter λ_j self-adaptively according to the grayscale features of the current image. (3) Neighborhood size influences the segmentation significantly. Nodule edges which resulted from $r = 1$ are relatively fine, especially where nodules are adhered to blood vessels or lung walls, as shown in the 5th and 6th images; however oversegmentations

and wrong segmentations to tiny blood vessels are easy to occur, as shown in the 4th and 5th images. Comparatively, neighborhoods when $r = 2$ are larger and the difference between grayscales at edges and the neighborhood averages increases. Parts of segmentation results are close to manual segmentations, like the 2nd and 6th images. However, under-segmentations occur sometimes, as shown in the 5th and 8th images, and it is also possible to segment nodules that are tightly adhered, as the 3rd image.

3.2. Executive Efficiencies of Algorithms. In order to further demonstrate the performance of the algorithm proposed by this paper, Table 5 shows average numbers of iterations (maximum number of iterations $L_{\max} = 100$) and running time (seconds) to obtain segmentation results in Figures 5–8 by the above five typical algorithms and our algorithm.

Table 5 indicates that when the FCM algorithm is employed to segment images, the more the iterations are needed to satisfy the convergence threshold value ϵ , the longer the program runs. Other algorithms (FCM.S, EnFCM, FGFCM, and FLICM) can converge with less iterations and have higher efficiencies. The algorithm proposed in this paper needs the least number of iterations and has the shortest running time, which is consistent with the observation described in Section 2.3 that fuzzy memberships change considerably after one iteration. In the meantime, prior knowledge plays a role in supervising unlabeled or sparsely labeled data, which speeds up the convergence of the objective function when improving the effect of clustering.

3.3. Error Rate of Segmentation. Figure 10 demonstrates the comparison between boundaries of all types of pulmonary nodules by our algorithm e_a and those obtained by manual segmentations of experts e_m . Figures 10(a), 10(b), 10(c), and 10(d) correspond to four types of pulmonary nodules in test samples with different amounts. In each plot of Figure 10, the ordinate represents number of pulmonary nodules and

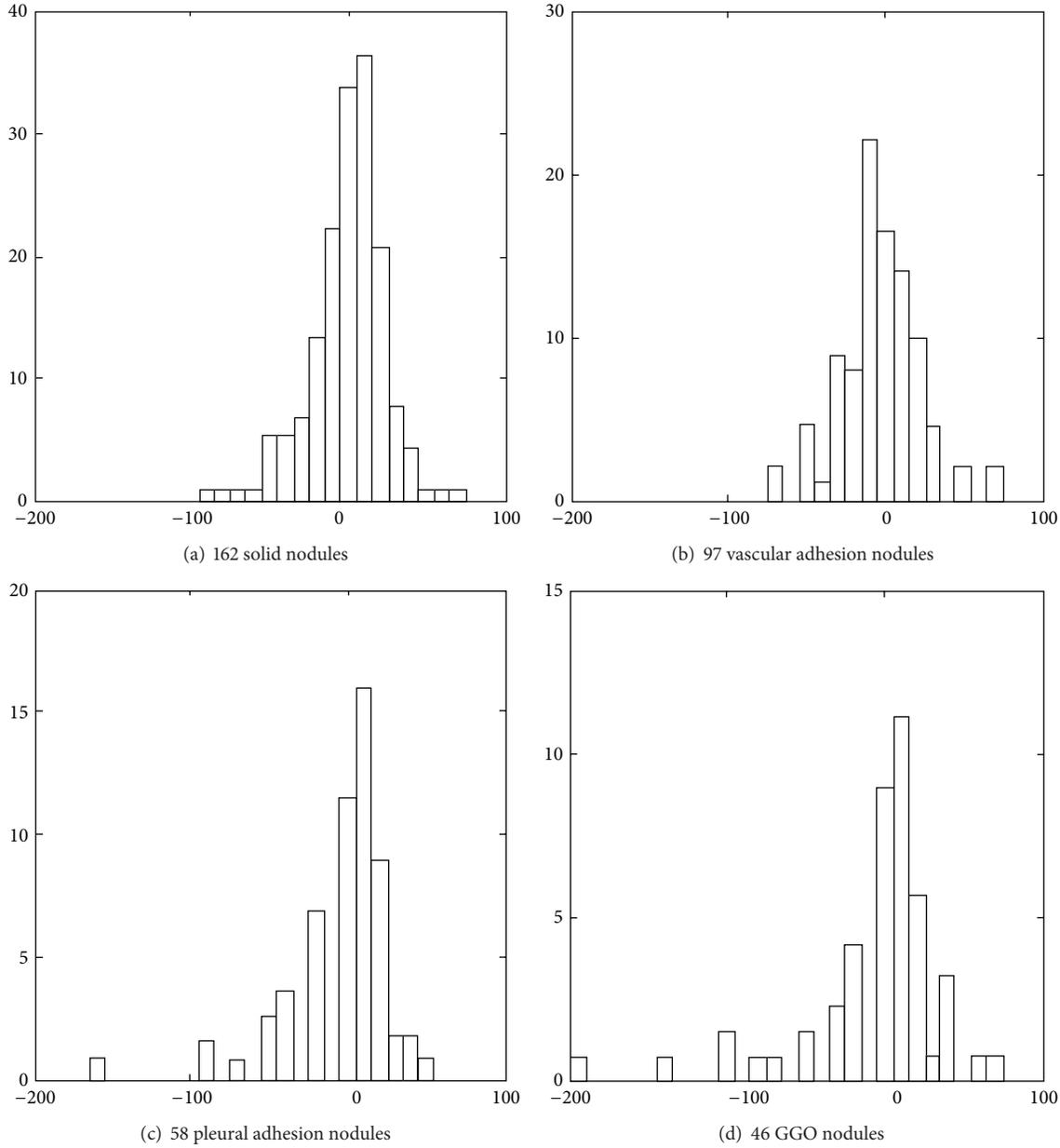


FIGURE 10: Histogram for comparison between our method and manual segmentation.

the abscissa represents rate of wrong segmentations $(e_m - e_a)/e_m$ (%). A positive rate of wrong segmentations means undersegmentation resulting from the algorithm with respect to manual segmentations. Otherwise, a negative rate of wrong segmentations means oversegmentation resulting from the algorithm with respect to manual segmentations.

It is seen from Figure 10 that the algorithm performs the worst for GGO nodules. The proportion of GGO nodules whose rates of wrong segmentations lie in the interval $[-50\%, +50\%]$ is less than 70%. This proportion for any of the other three types of nodules reaches 85%. It is also noticeable that the rates of wrong segmentations for some nodules approach -200 or 100 .

4. Conclusions

In this paper, a fast and self-adaptive FCM pulmonary nodule segmentation method combining clustering and classification learning is proposed in order to obtain satisfactory segmentations for blood vessel adhesion, pleura adhesion, and GGO pulmonary nodules. The new method improves the traditional FCM algorithm according to features of the above pulmonary nodules. It updates fuzzy memberships of central pixels based on grayscale similarities and spatial similarities of neighborhoods, so that clustering centers can be obtained quickly. The new method solves the problems of traditional segmentation methods such as the reliance on

the contrast between objectives and backgrounds, difficulty to obtain weak edges of pulmonary nodules and attached tissues, slow convergence of objective functions, and so forth. In the meantime, according to the artificial labeled data, we can realize the aided supervision for pulmonary nodule segmentation missing classification information. Experimental results indicate that the newly proposed algorithm can segment blood vessel adhesion, pleura adhesion, and GGO pulmonary nodules fast and exactly and performs better than traditional FCM, FCM.S, EnFCM, FGFCM, and FLICM in segmentation effects, executive efficiencies, and rates of error.

However, there are still some disadvantages about the new method. For example, it does not work well for segmenting tiny nodules (diameter < 1 cm) and segmentations of GGO nodules and tightly adhered nodules are unstable. In future research, we will solve these problems by extracting more characteristics of pulmonary nodules and deep learning method. We will also analyze the symptoms of benign and malignant nodules comprehensively and provide real aided diagnosis for early detection and screening of lung cancer.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] M. Puderbach and H. U. Kauczor, "Can lung MR replace lung CT?" *Pediatric Radiology*, vol. 38, no. 3, pp. S439–S451, 2008.
- [2] K. Doi, "Computer-aided diagnosis in medical imaging: historical review, current status and future potential," *Computerized Medical Imaging and Graphics*, vol. 31, no. 4-5, pp. 198–211, 2007.
- [3] J. Li, J. Dang, and Y. Wang, "Medical image segmentation algorithm based on quantum clonal evolution and two-dimensional tsallis entropy," *Journal of Computer-Aided Design and Computer Graphics*, vol. 26, no. 3, pp. 465–471, 2014.
- [4] X. Shen, K. Wang, and Q. Guo, "Local thresholding with adaptive window shrinkage in the contourlet domain for image denoising," *Science China Information Sciences*, vol. 56, no. 9, pp. 1–9, 2013.
- [5] A. P. Reeves, A. B. Chan, D. F. Yankelevitz, C. I. Henschke, B. Kressler, and W. J. Kostis, "On measuring the change in size of pulmonary nodules," *IEEE Transactions on Medical Imaging*, vol. 25, no. 4, pp. 435–450, 2006.
- [6] S. Magalhães Barros Netto, A. Corrêa Silva, R. Acatauassú Nunes, and M. Gattass, "Automatic segmentation of lung nodules with growing neural gas and support vector machine," *Computers in Biology and Medicine*, vol. 42, no. 11, pp. 1110–1121, 2012.
- [7] F. Dong and J. Peng, "Brain MR image segmentation based on local Gaussian mixture model and nonlocal spatial regularization," *Journal of Visual Communication and Image Representation*, vol. 25, no. 5, pp. 827–839, 2014.
- [8] C. McIntosh and G. Hamarneh, "Medial-based deformable models in nonconvex shape-spaces for medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 31, no. 1, pp. 33–50, 2012.
- [9] K. Murphy, B. van Ginneken, A. M. R. Schilham, B. J. de Hoop, H. A. Gietema, and M. Prokop, "A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification," *Medical Image Analysis*, vol. 13, no. 5, pp. 757–770, 2009.
- [10] A. Ortiz, J. M. Gorrioz, J. Ramirez, and D. Salas-Gonzalez, "Improving MR brain image segmentation using self-organising maps and entropy-gradient clustering," *Information Sciences*, vol. 262, pp. 117–136, 2014.
- [11] X. Zhang, C. Zhang, W. Tang, and Z. Wei, "Medical image segmentation using improved FCM," *Science China Information Sciences*, vol. 55, no. 5, pp. 1052–1061, 2012.
- [12] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient N-D image segmentation," *International Journal of Computer Vision*, vol. 70, no. 2, pp. 109–131, 2006.
- [13] X. Chen, J. K. Udupa, U. Bagci, Y. Zhuge, and J. Yao, "Medical image segmentation by combining graph cuts and oriented active appearance models," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2035–2046, 2012.
- [14] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, NY, USA, 1981.
- [15] K. S. Chuang, H. L. Tzeng, S. Chen, J. Wu, and T.-J. Chen, "Fuzzy c-means clustering with spatial information for image segmentation," *Computerized Medical Imaging and Graphics*, vol. 30, no. 1, pp. 9–15, 2006.
- [16] M. N. Ahmed, S. M. Yamany, N. Mohamed, A. A. Farag, and T. Moriarty, "A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data," *IEEE Transactions on Medical Imaging*, vol. 21, no. 3, pp. 193–199, 2002.
- [17] M. A. Balafar, A. R. Ramli, S. Mashohor, and A. Farzan, "Compare different spatial based fuzzy-c-mean (FCM) extensions for MRI image segmentation," in *Proceedings of the 2nd International Conference on Computer and Automation Engineering (ICCAE '10)*, vol. 5, pp. 609–611, IEEE, Singapore, February 2010.
- [18] S. Chen and D. Zhang, "Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 34, no. 4, pp. 1907–1916, 2004.
- [19] L. Szilagyi, Z. Benyo, S. M. Szilagyi, and H. S. Adam, "MR brain image segmentation using an enhanced fuzzy C-means algorithm," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 724–726, September 2003.
- [20] W. Cai, S. Chen, and D. Zhang, "Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation," *Pattern Recognition*, vol. 40, no. 3, pp. 825–838, 2007.
- [21] S. Krinidis and V. Chatzis, "A robust fuzzy local information C-means clustering algorithm," *IEEE Transactions on Image Processing*, vol. 19, no. 5, pp. 1328–1337, 2010.
- [22] J. C. Bezdek, "A physical interpretation of fuzzy isodata," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 6, no. 5, pp. 387–389, 1976.
- [23] S. G. Armato III, G. McLennan, L. Bidaut et al., "The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans," *Medical Physics*, vol. 38, no. 2, pp. 915–931, 2011.