

Received December 20, 2019, accepted January 16, 2020, date of publication February 3, 2020, date of current version March 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2968936

Segmentation of Pulmonary Nodules Using a GMM Fuzzy C-Means Algorithm

XIANGXIA LI¹, BIN LI², FANG LIU³, HUA YIN¹, AND FENG ZHOU¹

¹School of Information Engineering, Guangdong University of Finance and Economics, Guangzhou 510320, China

²School of Automation Science and Engineering, South China University of Technology, Guangzhou 510641, China

³School of Internet Finance and Information Engineering, Guangdong University of Finance, Guangzhou 510521, China

Corresponding author: Fang Liu (1142194635@qq.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61305038, Grant 61273249, and Grant 11901113, in part by the Guangdong Youth Innovation Talent Project, China, (Natural Sciences) under Grant 2018KQNCX086, Grant 2018KQNCX170, and Grant 2017KQNCX083, in part by the Public Science and Technology Research Funds Projects of Ocean under Grant 201505002, in part by the Fundamental Research Funds for the Central Universities, (SCUT), under Grant 2015ZZ028, in part by the Guangzhou Science and Technology Plan Project, China, under Grant 201904010225, and in part by the Guangzhou Science and Technology Project of China under Grant 201707010495.

ABSTRACT Segmentation of pulmonary nodule in thoracic computed tomography (CT) plays an important role in the computer-aided diagnosis (CAD) and clinical practices. However, segmentation of pulmonary nodules still remains a challenging task due to the presence of intrinsic noise, low contrast, intensity-profile inhomogeneity, variable sizes and shapes. Many variants and extensions of fuzzy C-mean (FCM) clustering algorithm have been developed to preserve image details as well as suppress image noises. However, these variants overemphasize the importance of the spatial information and neglect the role of the prior knowledge. To address this problem, a GMM fuzzy C-means (GMMFCM) algorithm is proposed for the segmentation of pulmonary nodules in this paper. A novel local similarity measure is defined by using local spatial information and GMM statistical information. A neighboring term is added to the energy function of traditional fuzzy C-mean algorithm. A superpixel-based random walker is proposed to segment pulmonary parenchyma, which reduces the computational complexity and improves the segmentation performance. Experiments performed on the LIDC dataset and the GHGZMCPLA dataset demonstrate that the segmentation performance of proposed GMMFCM algorithm is superior to the state-of-the-art algorithms.

INDEX TERMS Segmentation, lung cancer, pulmonary nodule, random walker, FCM, Gaussian mixture model.

I. INTRODUCTION

Lung cancer is the leading cause of cancer-related death worldwide. The American Cancer Society estimates that 1,688,780 new cancer cases and 600,920 cancer deaths will be diagnosed in the United States in 2017. The five-year survival rate of lung cancer patients is only 18%, because most lung cancer patients are diagnosed at an advanced stages [1]. Computed tomography (CT) is one of the most common imaging modalities for examining and screening lung cancer. Several studies have proved that early diagnosis of lung cancer could help improve survival rate and reduce the mortality rate by up to 20% [1]. Lung cancer potentially manifests itself as pulmonary nodules in an early stage. Although most

pulmonary nodules with a size equal to or less than 5 mm are benign, distinguishing lung cancer from benign lesions is crucial. Therefore, an early diagnosis of pulmonary nodules is essential to determine whether the treatment is necessary.

The accurate segmentation of pulmonary nodules in CT images is an important task for the early diagnosis of lung cancer. Traditionally, the pulmonary nodules are segmented by the radiologists. However, manual segmentation is time-consuming and prone to intra- and inter-observer variability, and thus making the results unreliable. In addition, the vast amount of CT images to be analyzed bring an invisible burden for radiologists. Hence, automatic segmentation of pulmonary nodules in the thoracic CT images is an area of ongoing and extensive research. Clinicians burdened in the situations, which occur very often in clinical practices, will indeed benefit from the automated segmentation methods.

The associate editor coordinating the review of this manuscript and approving it for publication was Gina Tourassi.

Segmentation of pulmonary nodules still remains a challenging task due to intra-nodules inhomogeneity, low contrast, blurred boundaries, variable sizes and irregular shapes. During the last decades, plenty of efforts have been devoted to pulmonary nodule segmentation in the literature, such as threshold-based methods [2], region growth-based methods [3], morphology-based methods [4], [5], shape-based methods [6], Hessian-based methods [7], [8], active contour-based methods [9]–[11], level set-based methods [12], [13], clustering-based methods [14], [15], local shape analysis methods [16] and machine learning-based methods [17], [18]. Threshold-based and region growth-based methods can achieve good segmentation performance for solitary pulmonary nodules, because intensity distributions are distinguishable between the nodule and background. The segmentation of other pulmonary nodules often fails due to the strong overlap of respective intensity distributions, such as juxta-vascularized nodule, juxta-pleural, and ground-glass opacity (GGO). Shape-based methods have been well studied for pulmonary nodule segmentation. They model a nodule either as a spherical or ellipsoidal shape. However, due to the large morphology variability of pulmonary nodules, their geometrical assumptions are not valid. Active contour and level set methods have been attracting more and more attentions for pulmonary nodule segmentation. Although these deformable methods have obtained the satisfactory segmentation performances, they are sensitive to noise and depend on the location of initial contour. In addition, due to the use of a huge number of iterations, these deformable methods demand the high computational cost. Therefore, it is of great interest to develop a simple and general algorithm for pulmonary nodule segmentation.

Fuzzy C-means (FCM) algorithm has been widely used in image segmentation. The main drawback of the conventional FCM algorithm is sensitive to noise due to the lack of the spatial information. However, the presence of intensity inhomogeneity and noise is inevitable in the CT images. Although the variations of FCM algorithm yield the desirable results for image segmentation, there still exist some shortcomings. For example, some variations suffer from the limitation of segmentation accuracy. Another variations can suffer from the problem of noise and outliers. Selecting the initial cluster centers is also a challenging problem.

These drawbacks significantly hinder the applications of FCM algorithm on pulmonary nodule segmentation. To address these problems, a GMM fuzzy C-mean (GMMFCM) algorithm is proposed for pulmonary nodules segmentation in this paper. Distinct from the FLICM algorithm, the proposed algorithm incorporates GMM prior knowledge into the traditional FCM algorithm by defining a new local similarity measure.

The contributions of this paper are as follows:

1) A novel multiscale dot enhancement filter is proposed. The traditional multiscale dot enhancement filter is sensitive to image noise due to the calculation of second order derivatives of Hessian matrix. To address

this problem, a novel multiscale dot enhancement filter is proposed by incorporating the Hessian matrix and shape index (SI) to alleviate the interference of the intensity inhomogeneity within pulmonary nodules, as well as avoid the influence of image noise and surrounding tissues.

2) The seeds of pulmonary nodule and background are automatically selected. How to choose the proper seeds is very important for the likelihood estimation of GMM model. In this paper, the nodule seeds are chosen from the response results of the multiscale dot filter, and the background seeds are chosen by jointing the shape index and texture properties.

3) A novel local similarity is proposed by using GMM prior knowledge. The traditional FLICM algorithm fails to discriminate the nodule pixels and non-pixels in CT images, resulting in non-optimal segmentation results. This severely restricts the application of FLICM algorithm. To overcome this limitation, a novel local similarity measure is proposed by using GMM based on posterior probability. The posterior probability is calculated according to nodule and background seeds.

4) The initial cluster centers selection. For the previous FCM algorithms, the initial cluster centers were selected randomly. In this paper, the initial cluster centers of the nodules are chosen from the multiscale dot filter, and the initial cluster centers of background are chosen from background seeds.

The remainder of this paper is organized as follows: Section II gives the previous studies related to our work. Section III presents a detailed description of the proposed GMMFCM algorithm for segmentation of pulmonary nodules. Section IV describes the results of pulmonary nodule segmentation, as well as the comparison of the proposed with other related segmentation algorithms on the LIDC dataset and the GHGZMCPLA dataset. Section V gives the discussions of this paper. Finally, section VI provides the conclusions and future work.

II. RELATED WORKS

Fuzzy C-Means (FCM) clustering is one of the most popular clustering algorithms. FCM is a soft clustering algorithm in which each image pixel can belong to more than one cluster with different degrees of membership. FCM has been successfully applied in several image processing problems due to its simplicity and efficiency, such as image clustering [19]–[21], image segmentation [22]–[26] and image classification [27]–[32]. The traditional FCM algorithm achieves the segmentation results by minimizing the sum of the distance between the pixel and its corresponding cluster centers. Although it works well on intensity homogeneous and noise-free images, it fails to segment images corrupted by noise and other imaging artifacts. It is because it does not consider the spatial information and uses the square of Euclidean distance. However, medical images are often corrupted by noise and the artifacts due to the physical mechanisms of the acquisition process and the movement of patient's body. Many variants and extensions have been developed in the literature. Many researchers took the local spatial information

into account. For example, Huang *et al.* [23] proposed an FCM_S algorithm for segmentation of brain MRI images. The authors introduced a spatial neighborhood term into the original objective function of FCM algorithm. The main shortcoming of FCM_S was its computational complexity and storage requirement, because the neighborhood labels of pixels were calculated in each iteration step. Its variants FCM_S1 and FCM_S2 algorithms were developed to reduce the computational burden of the spatial neighborhood term. To further speed up the convergence rate, Cai *et al.* [33] proposed fast generalized FCM (FGFCM) for image segmentation. The authors combined the spatial and gray levels information to construct a local similarity measure, then formed a non-linearly weighted sum image. The clustering process was performed on the summed image rather than the original image. Unfortunately, some important information of the original image may be lost, thus resulting in degrading image segmentation performance.

Another critical issue is how to determine the crucial parameters, which controls the tradeoff between robustness of noise and effectiveness of preserving image details. However, it is very difficult to select appropriate parameters. To address these problems, Krinidis *et al.* [34] proposed a FLICM algorithm for image segmentation. The authors introduced a fuzzy factor to replace the parameters used in FCM_S algorithm and its variants. Although FLICM algorithm achieved the desired segmentation results, there are still some shortcomings in keeping the image edge. Li *et al.* [35] proposed a modified FLICM algorithm for image segmentation, which introduced edge and local information to reduce edge degradation. Gong *et al.* [36] proposed a variant of FLICM algorithm, which introduced a tradeoff weighted fuzzy factor and kernel metric for image segmentation. Although FLICM algorithm and its variants considered the influence of the neighboring pixels, they failed to make full use of the information of the center pixel in the local window. Thus, the borders and edges of some regions may be over-smoothed. To reduce the smoothness and further improve the segmentation accuracy, Ji *et al.* [37] proposed a RSCFCM algorithm for brain MR image segmentation. The authors introduced a spatial factor based on posterior probabilities and prior probabilities to alleviate the disturbance of noise and intensity inhomogeneity. The dissimilarity function was constructed by considering the prior probabilities. Recently, Zhang *et al.* [38] proposed an ADFLICM algorithm by introducing a fuzzy local similarity measure to replace the fixed parameter of FCM_S algorithm. The main advantage of the fuzzy local similarity measure is that it can adaptively vary by the gray level and local spatial relationships between the center pixel and its neighbors in a local window. Zhao *et al.* [39] proposed interval type-2 fuzzy C-Mean clustering algorithm for color image segmentation. A novel interval type-2 fuzzy clustering objective function is constructed by utilizing the intuitionistic fuzzy information extracted from images. Singh and Bala [40] proposed the local and nonlocal FCM, the distance function of the FCM is represented as the sum of the local and nonlocal distances

which themselves are the weighted values of the Euclidean distance used in the FCM.

III. METHODOLOGY

In this paper, a GMM fuzzy C-means clustering algorithm is proposed for pulmonary nodule segmentation. The flowchart of the proposed method is shown in Fig. 1. First, the original CT thoracic images are smoothed by non-local mean filter and down-sampled by Gaussian pyramid. Second, the pulmonary parenchyma is segmented by superpixel-based random walker algorithm. To generate the nodule seeds, pulmonary nodules are enhanced by a new multiscale dot enhancement filter. To generate background seeds, a scheme is proposed by jointing shape index and texture features. Then, the nodule and background GMM models are built by generating the seeds. Finally, the pulmonary nodules are segmented by GMM fuzzy C-mean algorithm. The details of each step are described in the following sections.

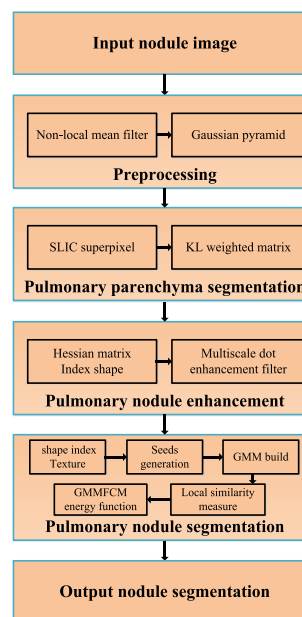


FIGURE 1. The flowchart of the proposed algorithm.

A. PREPROCESSING

Due to the imaging mechanism of acquisition process and the movement of the patient body, thoracic CT images are inevitably contaminated by the noise and artifacts. Therefore, the preprocessing is necessary to reduce the noise and enhance the contrast.

Firstly, a non-local mean (NLM) filter with a mask of 3×3 is adapted in this section. The non-local mean (NLM) filter is capable of reducing the noise without sacrificing the details of the image. Secondly, the Gaussian pyramid is employed to reduce the image resolution by half, which reduces the computational complexity and accelerates the convergence rate. The pyramid is a sequence of smoothed images generated by a Gaussian filter. The basis of the Gaussian pyramid is

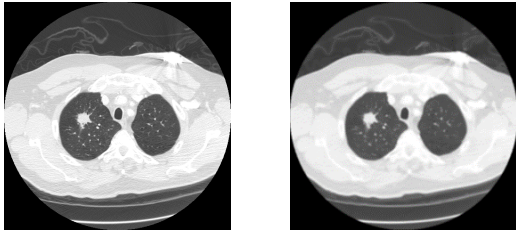


FIGURE 2. An Example of CT image with a pulmonary nodule in a preprocessing step. (a) Original CT images with a well-circumscribed nodule. (b) Corresponding subsampled image obtained by using Gaussian pyramid.

formulated in equation (1).

$$g_l(i, j) = \sum_{m=-2}^2 \sum_{n=-2}^2 w(m, n)g_{l-1}(2i + m, 2j + n) \quad (1)$$

where l is a level of the Gaussian pyramid. i and j are the spatial coordinate positions of x - and y -direction at the l -th level, respectively. g_{l-1} is the Gaussian smoothed image at the $l - 1$ -th level. $w(m, n)$ denotes a 5×5 Gaussian filter, which is used to the neighboring pixels of a pixel (i, j) . g_0 denotes an initial image, i.e. an input image at the level $l = 0$. g_l is generated by smoothing the image g_{l-1} with a Gaussian filter $w(m, n)$. Fig. 2 shows a CT image with a nodule. Fig. 2(a) is an original image with a well-circumscribed nodule. Fig. 2(b) is the down-sampled result by Gaussian pyramid. The image size of 512×512 is down-sampled to the size of 256×256 . This image is directly applied to segment pulmonary parenchyma process.

B. GMM FUZZY C-MEANS

The traditional FCM algorithm neglects the spatial information, hence it is more sensitive to noise. Although the FLICM algorithm introduced a fuzzy factor to alleviate the image noise and preserve the image details, it has the limitation in discriminating the boundary pixels. To address this problem, a GMM fuzzy C-mean algorithm (GMMFCM) is proposed for pulmonary nodule segmentation. The proposed GMMFCM takes into account the spatial relationship between the neighboring pixels, and computes the posterior probability of the pixels belonging to each cluster. A new local similarity measure is defined based on the posterior probabilities of the GMM models. Then, the objective function of the GMMFCM algorithm is defined. The details of each step are described in the next sections.

1) DEFINITION OF LOCAL SIMILARITY MEASURE

The objective function of FCM algorithm is iteratively minimized to find a solution to the problem. The high membership degree is assigned to the pixel whose intensity value is close to the center of the corresponding cluster, and the low membership degree is assigned to the pixel whose intensity value is far from the center. Hence, the membership degree is sensitive to the presence of noise and intensity inhomogeneity.

The FLICM algorithm introduced a fuzzy factor to reduce the image noise and simultaneously preserve image details. The fuzzy factor characterizes the spatial relationships between the pixel and its neighbors. However, it is unable to handle the boundary pixels. It is well-known that some pulmonary nodules and adjacent structures share the similar intensities, thus resulting in the fuzzy boundaries. To address this problem, a novel local similarity measure is introduced for pulmonary nodule segmentation by incorporating the local spatial information and the GMM statistics information. For two pixels i and j , the spatial measure is defined in equation (2).

$$SM_{ij}^{GMM}(k) = \frac{P_{ki} \times P_{kj}}{d_{ij}^2}, \quad k \in \{1, \dots, c\} \quad (2)$$

where $d(\cdot)$ is a spatial distance between two neighboring pixels i and j . P_{ki} and P_{kj} denote the posterior probabilities of the pixels i and j by fitting the Gaussian Mixture Model (GMM) of the cluster k . d_{ij} is the spatial Euclidean distance between the pixels i and j . c is the number of clusters.

A novel local similarity measure is defined in equation (3).

$$w_{ir}^{GMM} = \begin{cases} SM_{ir}^{GMM} & i \neq r \\ 0 & i = r \end{cases} \quad (3)$$

where the i -th pixel is the center of local window N_i , and the r -th pixel is the neighborhood pixel that falls into N_i . The proposed local similarity function does not involve any adjusted parameters. After the definition of local similarity measure, the objective function of GMMFCM algorithm is defined in the following section.

2) FORMULATION OF PROPOSED GMMFCM ALGORITHM

The objective function of the proposed GMMFCM algorithm is defined in equation (4).

$$J_{GMMFCM} = \sum_{i=1}^N \sum_{k=1}^c u_{ki}^m \|x_i - v_k\|^2 + \frac{\alpha}{N_R} \sum_{i=1}^N \sum_{k=1}^c u_{ki}^m \left(\left(1 - w_{ir}^{GMM} \right) \sum_{x_r \in N_i} \|x_r - v_k\|^2 \right) \quad (4)$$

s.t. $\sum_{k=1}^c u_{ki} = 1 \quad 0 \leq u_{ki} \leq 1 \quad \forall k, i$

where u_{ki} is the degree of membership of the i -th pixel belonging to the k -th cluster. v_k is the prototype of the k -th cluster. m is a fuzzy index, which determines the level of cluster fuzziness of the membership grades. N_R is the cardinality of the neighborhood system N_i . $\|\cdot\|$ denotes the norm operator. α is the tradeoff parameters, which controls the effect of the factor terms. The first term is the objective function used in traditional FCM algorithm, which assigns a high membership to the image pixels whose intensity values are close to the center of the particular cluster and a low membership to the image pixels whose intensity values are far from the center of the particular cluster. The second term is a local similar

regularizer term, whose weight factor is automatically determined by incorporates the spatial relationship and the prior knowledge information of two neighboring pixels. Therefore, local spatial relationship was adaptively changed and more local spatial information was considered.

According to the Lagrange multiplier method, the Lagrange function of (4) is defined in equation (5).

$$\begin{aligned}
 F_{GMMFCM} = & \sum_{i=1}^N \sum_{k=1}^c u_{ki}^m \|x_i - v_k\|^2 \\
 & + \frac{\alpha}{N_R} \sum_{i=1}^N \sum_{k=1}^c u_{ki}^m \left(\sum_{x_r \in N_i} (1 - w_{ir}^{GMM}) \|x_r - v_k\|^2 \right) \\
 & + \lambda \sum_{i=1}^N \left(1 - \sum_{k=1}^c u_{ki} \right) \quad (5)
 \end{aligned}$$

By the partial derivative of the Lagrange function (5) with respect to the variances u_{ki} , v_k , and setting them to be zero, the membership u_{ki} and the center v_k are updated according to the updating equations (6) and (7).

$$v_k = \frac{\sum_{i=1}^N u_{ki}^m \left(x_i + \frac{\alpha}{N_R} \sum_{x_r \in N_i} (1 - w_{ir}^{GMM}) x_r \right)}{\left(1 + \frac{\alpha}{N_R} \sum_{x_r \in N} (1 - w_{ir}^{GMM}) \right) \cdot \sum_{i=1}^N u_{ki}^m} \quad (6)$$

$$u_{ki} = \frac{1}{\sum_{j=1}^c \left(\frac{\|x_i - v_k\|^2 + \frac{\alpha}{N_R} \sum_{x_r \in N_i} (1 - w_{ir}^{GMM}) \|x_r - v_k\|^2}{\|x_i - v_j\|^2 + \frac{\alpha}{N_R} \sum_{x_r \in N_i} (1 - w_{ir}^{GMM}) \|x_r - v_j\|^2} \right)^{\frac{1}{m-1}}} \quad (7)$$

After cluster centers are updated, the stopping criterion is used to verify whether cluster centers have been converged or not. If cluster centers have been converged, then the algorithm is applied again to obtain the new fuzzy partitions and cluster centers. This process is repeated until the cluster centers are converged or the maximum number of iterations is reached. The GMMFCM algorithm is summarized in Algorithm 1.

3) BUILD OF NODULE AND BACKGROUND GMM MODELS

To calculate the posterior probabilities of local similarity function, the GMM models of the nodule and background are built by using the generated seeds in this section. To reduce the computational complexity, pulmonary parenchyma is segmented by the proposed superpixel-based random walker algorithm. Then, the nodule and background seeds are automatically generated by the proposed multiscale dot filter and a scheme of background seed generation.

In the case of normal lung, air-filled pulmonary parenchyma and surrounding tissues have a large density difference in the CT images. The simple thresholding methods are often used to segment the pulmonary parenchyma. However, segmentation of pulmonary parenchyma with high-density pulmonary nodules is a nontrivial problem. This is

Algorithm 1 GMMFCM

Input:

The number of cluster c , the stopping condition ε , fuzzy index m .

Output:

Clusters v_1, v_2, \dots, v_c and the new U .

- 1: Initial the fuzzy partition matrix U and the center vector V ;
- 2: Set the loop counter $b = 0$;
- 3: Calculate the new local similarity measure w_{ir}^{GMM} by equation (3);
- 4: Update the cluster center V by using equation (6);
- 5: Update the fuzzy partition matrix U by using equation (7);
- 6: Repeat the steps 3, 4 and 5, until the stopping criterion is satisfied.

$$\max_i \left| v_k^b - v_k^{(b+1)} \right| < \varepsilon, \quad k = 1, 2, \dots, c$$

Then, the iteration will stop; otherwise, let $b = b + 1$ and go back to the step 3 and repeat.

7: When the algorithm converges, a new fuzzy partition matrix U is obtained. Then, the optimization process can assign a pixel i to the cluster C with the largest membership value.

7: When the algorithm converges, a new fuzzy partition matrix U is obtained. Then, the optimization process can assign a pixel i to the cluster C with the largest membership value.

$$C_i = \arg \left\{ \max_k \{u_{ki}\} \right\}, \quad k = 1, 2, \dots, c$$

because lungs with pulmonary nodules may have density values similar to other anatomical tissues surrounding the lung regions. Only a few literature have been published that handle segmentation of lungs with pulmonary nodules [41], [42]. None of the existing segmentation methods directly segment lungs with pulmonary nodules at arbitrary locations.

Recently, random walker (RW) algorithm has become a hot topic in the field of computer vision. Regarding medical image segmentation, some researches have been aware of its advantages, and also have shown its benefits in segmentation of different organs and tissues [43]–[46]. The solution of the RW model is obtained by solving a linear system according to the norm used to define the energy function. So, the total time complexity depends on the number of pixels. Therefore, the computational complexity and memory cost can be enormous for a large image. To address this problem, the superpixel-based random walker (SRW) model is proposed for pulmonary parenchyma segmentation, which consists of two steps. The first step is to over-segment the CT image into some small regions via SLIC superpixel algorithm and the second step is to build RW model for pulmonary parenchyma segmentation.

Simple Linear Iterative Clustering (SLIC) superpixel algorithm [47] is employed to efficiently generate superpixels

due to its simplicity and efficiency. Each input image is over-segmented into a set of superpixels $\{R_i\}_{i=1}^M$ by SLIC superpixel algorithm, where M is the number of superpixels. Based on superpixels, an undirected weighted graph $G = (V, E)$ is constructed, where each superpixel is represented as a node and each edge connects two neighboring superpixels. Due to the great reduction of the number of image primitives, superpixel-level RW model can significantly reduce the computational complexity and memory requirement compared to pixel-level RW model. In addition, these superpixels provide compact and structural information, which can reduce the risk of assigning error labels to the corresponding pixels.

For each pixel within a superpixel, intensity, texture, gradient features and spatial distance are extracted. A multivariate normal distribution $N(u_i, \sum_i)$ of a superpixel is modeled in the eight dimensional feature space. Given two neighboring superpixels R_i and R_j , we measure the similarity of their probability distributions based on Kullback–Leibler (KL) divergence. The more similar two probability distributions are, the smaller the KL divergence is. The new weighted matrix is defined in equation (8).

$$w^{KL}(R_i, R_j) = 1 - (KL(f_i, f_j) + KL(f_j, f_i)) / 2 \quad (8)$$

where f_i and f_j are the multivariate Gaussians with their distribution $N(u_i, \sum_i)$ and $N(u_j, \sum_j)$. $KL(f_i, f_j)$ denotes the Kullback–Leibler (KL) divergence between two d -dimensional Gaussian distributions, which is calculated in equation (9).

$$KL(f_i, f_j) = \frac{1}{2} \left[(u_i - u_j)^T \sum_j^{-1} (u_i - u_j) + Tr \left(\sum_j^{-1} \sum_i \right) + \log \frac{|\sum_j|}{|\sum_i|} - d \right] \quad (9)$$

By minimizing the energy function of SRW model, the probability of a superpixel R_i is close to that of its neighboring superpixel R_j when a high weight $w^{KL}(R_i, R_j)$ is given. That is, two neighboring superpixels with similar probability distributions can get similar probabilities. Fig. 3 shows the segmentation results for a juxta-pleural nodule. Fig. 3(b) and (c) show results by using the conventional RW (red contour) and the proposed SRW algorithm (green contour) for pulmonary parenchyma segmentation. The conventional RW algorithm is more prone to produce incorrect results, as shown Fig. 3(b). This is because the Juxta-pleural nodules may damage the pleural surfaces, resulting in disrupting the integrity of the boundary of the lung. Some local segmentation errors of the pulmonary parenchyma with pulmonary nodules are indicated by red arrows. As shown Fig. 3(c), the SRW algorithm achieves the desirable segmentation results compared with the conventional RW algorithm for pulmonary parenchyma segmentation. After pulmonary parenchyma segmentation, the nodule and background seeds are generated to build the GMM models. A detailed description is given in the next section.

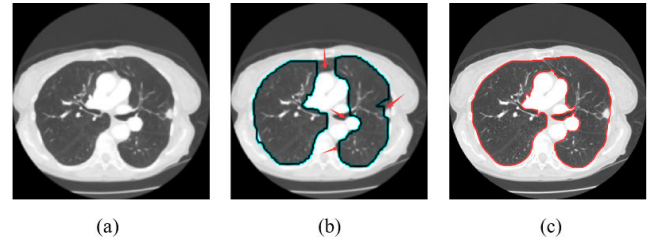


FIGURE 3. Segmentation results of pulmonary parenchyma with a juxta-pleural nodule. (a) The preprocessed CT image; (b) Segmentation results of the conventional RW algorithm. The red arrows indicate some local segmentation errors; (c) Segmentation results of the proposed the SRW algorithm.

To generate the nodule seeds, a new multiscale enhancement dot filter (NMEDF) is proposed in this section. The MEDF was firstly proposed by Li et al. [48] for pulmonary nodule enhancement. The MEDF can enhance dot-like pixels while suppressing other surrounding tissues, such as line-like vessels and planar-like airway walls. The traditional MEDF is highly sensitive to image noise, partial volume effects and patient motion. This is because the calculation of the Hessian matrix is based on the second-order partial derivatives. The novel MEDF incorporates Hessian matrix and the shape index [49], which is more robust against the intensity inhomogeneity within pulmonary nodules and image noise, as well as avoid the influence of surrounding tissues.

Let λ_1 and λ_2 denote two eigenvalues of the Hessian matrix in two-dimensional (2D) image space, which are ordered based on their magnitudes such as $|\lambda_2| \geq |\lambda_1|$. The shape index (SI) is calculated in equation (10).

$$SI(\sigma) = \frac{1}{2} - \frac{1}{\pi} \arctan \frac{\lambda_1 + \lambda_2}{\lambda_2 - \lambda_1} \quad (10)$$

Then, the likelihood function of the novel multiscale dot enhancement filter is defined in equation (11).

$$Z_{\sigma_i}^{dot}(\lambda_1, \lambda_2) = \begin{cases} g(SI(\sigma_i)) \cdot \frac{|\lambda_2|^2}{|\lambda_1|} & \lambda_2 = \lambda_1 < 0 \\ 0 & otherwise \end{cases} \quad (11)$$

where $g(SI(\sigma_i))$ is an indicator function such that $g(SI(\sigma_i))$ when the value of $SI(\sigma_i)$ is greater than or equal to 0.75 to find only dot-like shape. Otherwise, when the value of $SI(\sigma_i)$ is less than to 0.75, the value of $g(SI(\sigma_i))$ is equal to 0. If the sizes of pulmonary nodules with diameter range from d_0 to d_1 , then each scale factor is calculated in equation (12).

$$\sigma_1 = \frac{d_0}{4}, \quad \sigma_2 = r\sigma_1, \quad \sigma_3 = r^2\sigma_1, \dots, \quad \sigma_N = r^{N-1}\sigma_1 = \frac{d_1}{4} \quad (12)$$

where $r = (d_1/d_0)^{\frac{1}{N-1}}$ and N is the number of the scale factors. At each scale σ , the finally output of enhancement filter is calculated as the maximum response for a pixel x ,

which is represented in equation (13).

$$Z_{\max} = \max_{\sigma_{\min} \leq \sigma \leq \sigma_{\max}} \left(\sigma_i^2 \cdot Z_{\sigma_i}^{\text{dot}} \right), \quad i=1, 2, \dots, N \quad (13)$$

The nodule seeds are generated by thresholding the enhanced pulmonary nodule image. To generate the background seeds, shape index and texture features are used. The pixels with shape index map less than a threshold T_1 are extracted as the background seeds. However, the pixels may be mistakenly identified by shape index. To address this problem, the texture feature is considered. The Gabor transform is employed to generate the texture map. All pixels with texture map less than a threshold T_2 are extracted as the background seeds. The finally background seeds are obtained by jointing shape index and texture properties. After the seeds are obtained, the nodule and the background GMMs are built to calculate the posterior probability of local similarity function.

Fig. 4 shows the generation results of nodule and background. Fig. 4(a) is a segmented pulmonary parenchyma with a pulmonary nodule. Fig. 4(b) and (c) are the nodule enhancement results by the traditional NMEDF and the proposed NMEDF, respectively. As shown in Fig. 4(b) and (c), the proposed MEDF obtains a better nodule enhancement results and suppresses more background regions. The red arrows indicate the enhanced nodules. The good result benefits from the incorporation of Hessian matrix and shape index. After pulmonary nodule enhancement, a suspicious pulmonary nodule within the region of interest is obtained by threshold-based technique, which is shown in Fig. 4(d). Fig. 4(e) is the nodule seeds obtained by thresholding the enhanced image, which are spotted with red points. Fig. 4(f) and (g) are background seeds obtained by thresholding the shape index map and the texture map, respectively. As shown in Fig. 4(f), a small part

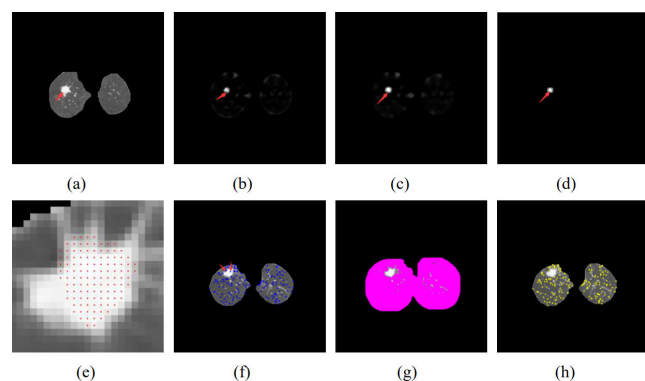


FIGURE 4. Generation of pulmonary nodule and background seeds. (a) The segmented pulmonary parenchyma; (b) Enhancement result of the conventional MEDF. (c) Enhancement result of the proposed MEDF; (d) The enhanced nodules within the region of interest obtained by threshold-based technique. (e) The nodule seeds obtained by thresholding the enhanced image, which are spotted with red points; (f) The background seeds obtained by thresholding the shape index map, which are spotted with blue points; (g) The background seeds obtained by thresholding the texture map, which are spotted with purple points; (h) The final background seeds obtained by the joint of shape index and texture properties, which are spotted with yellow points.

of the pixels within the pulmonary nodule region are regarded as the background seeds (red arrows) when the shape index is only used. As shown Fig. 4(h), the wrong background seeds is removed by jointing the shape index and texture properties.

IV. EXPERIMENTAL SETUP

In this section, the experiments are performed on two datasets to validate the performance of the proposed method. The segmentation performance of the proposed GMMFCM algorithm is evaluated by the qualitative and quantitative analysis. We compare the accuracy and effectiveness of the proposed algorithm with FCM, FCM_S, FCM_S1, FCM_S2, FLICM, FCM-type and Non-FCM algorithms. All experiments are run in the MATLAB platform on a PC with Intel E3-1225 CPU (3.31 GHz) and 4 GB RAM.

A. DATASETS

The LIDC dataset is the largest publicly available computed tomography (CT) images for validating the segmentation or classification performance of pulmonary nodules, which consists of 1018 cases from seven academic centers and eight medical imaging companies worldwide. Each subject contains two parts: the images from a clinical thoracic CT scan and the corresponding XML file that records the segmentation results of two-phase image annotation process performed by four experienced chest radiologists. The diameters of pulmonary nodule range from 2.03 mm to 38.12 mm, and the intervals of the slices range from 0.45 mm to 5.0 mm. All pulmonary nodules are segmented by up to four radiologists. The coordinates of the outlines of the nodules ≥ 3 mm are individually marked by four radiologists. The nodules < 3 mm are represented by a single coordinate.

In this paper, the nodules ≥ 3 mm (a total of about 893 pulmonary nodules) are selected to conduct the experiments. The nodules < 3 mm are not considered. Due to the inter-variability of segmentation results among four radiologists, a 50% consensus criterion [4] is used to produce the outline of ground truth. For all 893 nodules, 200 nodules are randomly selected for performance evaluation in this paper.

B. EVALUATION METRICS

To quantitatively evaluate segmentation performance of the proposed GMMFCM algorithm, seven evaluation criteria are used in this paper, including **Accuracy**, **Sensitivity**, **Specificity**, **False positive ratio (FPR)** and **False negative ratio (FNR)**, **Overlap score** and **Dice similarity coefficient (DSC)**.

Accuracy is the proportion of correctly identified pixels in the image segmented by the algorithms. It is calculated in equation (14).

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \quad (14)$$

Sensitivity is the proportion of correctly identified as nodule pixels and **specificity** is the proportion of correctly identified as background pixels, which are calculated in

equation (15) and (16).

$$Se = \frac{TP}{TP + FN} \tag{15}$$

$$Sp = \frac{TN}{TN + FP} \tag{16}$$

FPR is the fraction of falsely identified as nodule pixels and false negative ratio (FNR) is the fraction of falsely identified as background pixels, which are defined in equation (17) and (18).

$$FPR = \frac{FP}{TP + TN} \tag{17}$$

$$FNR = \frac{FN}{TP + TN} \tag{18}$$

Overlap score is a similarity measure, which reflects how the segmentation result of the algorithms matches the ground truth. It is calculated in equation (19).

$$Overlap = \frac{TP}{TP + FP + FN} \tag{19}$$

DSC is an overlap measure between the segmented pulmonary nodule and ground truth. It is calculated in equation (20).

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{20}$$

where true positive (TP) is the number of correctly identified as nodule pixels. False positive (FP) is the number of wrongly identified as nodule pixels. True positive negative (TN) is the number of correctly identified as background pixels. False Negative (FN) is the number of wrongly identified as background pixels. The values of seven evaluation criteria range from 0 to 1. The larger accuracy, sensitivity, specificity, overlap score and DSC are, the higher similarity between the segmented nodule and ground truth is. The smaller FPR and FNR are, the better segmentation performance of pulmonary nodules is.

C. PARAMETER SRTTING

In this paper, the parameters of the proposed GMMFCM method were optimized in all experiments. In the pulmonary nodule enhancement stage, the parameter σ should enough large to cover all the size of nodules, which is set the range from 2.03 mm to 68.43 mm at steps of 0.5 for all the experiments. Hence, the small or large pulmonary nodules can be successfully enhanced. d_0 and d_1 are set as 4 and 36 mm in the implementation, respectively. In the pulmonary nodule segmentation stage, there are four parameters, namely, T_1 , T_2 , α . In the case of the generation of nodule and background seeds method, the thresholds T_1 and T_2 control the numbers of nodule and background seeds, which are set to 0.9 and 0.7 to identify the image pixels that belong to the background, respectively.

The parameter α controls the effect of the factor terms. If the parameter α is set to 0, the proposed GMMFCM algorithm will degenerate to the conventional FCM algorithm.

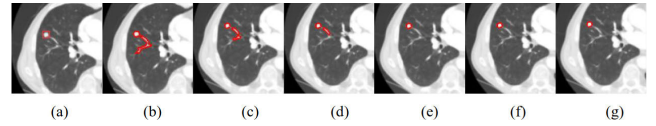


FIGURE 5. Segmentation results of the proposed algorithm with varying the parameter α . (a) the preprocessing image; (b) $\alpha = 0$; (c) $\alpha = 0.1$; (d) $\alpha = 0.2$; (e) $\alpha = 0.6$; (f) $\alpha = 0.8$; (g) $\alpha = 1$.

The segmentation performance of the proposed GMMFCM algorithm depends on the choice of parameters. To validate the dependence of segmentation accuracy on the regularization parameter α , a trial-and-error process is used to estimate the parameters. Firstly, the regularization parameter α were set the initial value $\alpha_{init} = 0$. Then, a range of parameter values are set to verify the performance of the proposed algorithm. Fig. 8 shows the segmentation results using the proposed algorithm with varying the parameter α . As shown in Fig. 5, the segmentation results slightly go down with the increasing the parameter α , which are shown in Fig. 5(b), (c) and (d). In a proper range, the proposed algorithm remains relatively stable for the regularization parameter, which are shown in Fig. 5(e), (f) and (g). After the several iterations, the segmentation results are desired when the parameters > 0.6 . Based on the above analysis it can be concluded, that the parameter are directly set to $\alpha = 0.8$ and in all experiments to obtain the best segmentation results. All experiments are performed on two datasets based on these optimal parameters.

D. QUALITATIVE RESULTS AND COMPARISONS

In this section, some qualitative comparisons are conducted on two datasets to evaluate the segmentation accuracy of the proposed algorithm.

1) QUALITATIVE EVALUATION ON THE LIDC DATASET

To qualitatively evaluate the proposed GMMFCM algorithm on the LIDC dataset, we apply the proposed algorithm to segment the CT slice with various types of pulmonary nodules and compare the segmentation results with ground truths. Four CT slice examples with different types of pulmonary nodules are obtained from the LIDC dataset for comprehensive validation of the proposed algorithm. The segmentation results are compared with that of the corresponding ground truths. The red and green contour indicate the segmentation results by the proposed algorithm and ground truth, respectively.

The visualization of the obtained segmentation results is shown in Fig. 6. The CT images after the preprocessing are shown in the leftmost column of Fig. 6. The red rectangles indicate the pulmonary nodule regions. The second column of Fig. 6 shows the close-ups of the segmentation results. From the second column of Fig. 6, we can see clearly that the results of the proposed algorithm is very close to the pulmonary nodule boundaries. As shown in the rightmost column of Fig. 6, the proposed GMMFCM algorithm produces

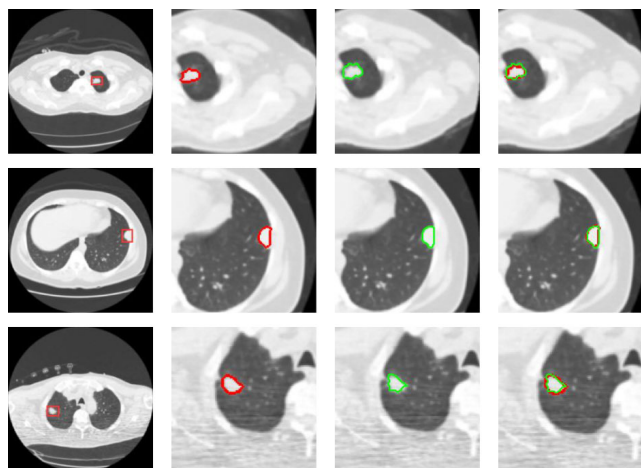


FIGURE 6. Examples of segmentation results on the LIDC dataset using GMMFCM algorithm for different types of pulmonary nodules. The first column: The CT images after the preprocessing; The second column: the close-ups of the segmentation results; The third column: the contours of the ground truth; The fourth column: the comparison between the segmentation results and ground truths.

the segmentation results closer to the ground truths. Although some pulmonary nodules are extremely subtle with fuzzy boundaries or homogeneous intensity, the proposed GMMFCM algorithm can successfully segment these nodules. The experiment results demonstrate that the proposed GMMFCM algorithm achieves the desired segmentation results for different types of pulmonary nodules from the LIDC dataset.

To qualitatively evaluate the superior segmentation results of the proposed GMMFCM algorithm, three comparison experiments are conducted on the LIDC dataset. The comparison experiments are performed by using five different algorithms (i.e., FCM, FCM_S, FCM_S1, FCM_S2, FLICM). Each algorithm directly uses the best parameters in the experiments and then we perform the experiments on three different types of pulmonary nodules, including a juxta-vascularized nodule, a juxta-pleural nodule and a GGO nodule.

Fig. 7 illustrates the segmentation results derived from the FCM, FCM_S, FCM_S1, FCM_S2, FLICM and GMMFCM algorithms for a juxta-vascularized pulmonary nodule, respectively. Fig. 7(a) is a lung CT slice with a juxta-vascularized pulmonary nodule. Fig. 7(b)-(f) are the segmentation contours obtained by using FCM, FCM_S, FCM_S1, FCM_S2, FLICM algorithms. As shown

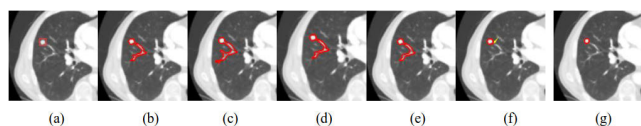


FIGURE 7. Comparison of the segmentation results using FCM, FCM_S, FCM_S1, FCM_S2, FLICM and algorithms for a CT image with a juxta-vascularized nodule. (a) A CT image with a juxta-vascularized nodule, marked with a red rectangular; (b) FCM algorithm; (c) FCM_S algorithm; (d) FCM_S1 algorithm; (e) FCM_S2 algorithm; (f) FLICM algorithm; (g) the proposed algorithm.

in Fig. 7(b), FCM algorithm produces the inferior segmentation result for the juxta-vascularized nodule. Many pixels belonging to the vessel structure are misclassified as a part of the pulmonary nodule. As shown in Fig. 7(c)-(e), the FCM_S, FCM_S1, FCM_S2 algorithms also produce the serious over-segmentation results for the juxta-vascularized nodule. Although they improve the traditional FCM algorithm to some extent by considering the spatial neighboring information, the negative effects of intensity similarity profiles between the pulmonary nodule and its adjacent vessel structure cannot be completely resolved. The FLICM algorithm obtains an improved segmentation result by introducing a fuzzy factor, which is shown Fig. 7(f). The most of the over-segmentation pixels are removed by the FLICM algorithm, but some pixels belonging to the vessel structure still remain, which are indicated by yellow arrows. This is because the FLICM algorithm has some shortcomings in identifying the pixels of weak boundary. Fig. 7(g) shows the segmentation contour of the proposed algorithm. As observed from Fig. 7(g), the proposed algorithm corrects the pixels who are misclassified as nodule pixels, and produces the desirable segmentation result. The proposed algorithm incorporates the spatial and statistical information to reduce boundary degradation.

Fig. 8 illustrates the segmentation results derived from the FCM, FCM_S, FCM_S1, FCM_S2, FLICM and GMMFCM algorithms for the juxta-pleural pulmonary nodule, respectively. Fig. 11(b)-(f) are the segmentation contours obtained by using FCM, FCM_S, FCM_S1, FCM_S2, FLICM algorithms. As shown in Fig. 8(b)-(e), the FCM, FCM_S, FCM_S1, FCM_S2 algorithms also produce the over-segmentation results for the juxta-pleural pulmonary nodule. From Fig. 8(f), we can clearly observe that the FLICM algorithm produces slightly over-segmentation result for juxta-pleural nodule, which is indicated by yellow arrows. The negative effects of intensity similarity profiles between the pulmonary nodule and its adjacent pleural structure also cannot be completely resolved. This is because the FLICM algorithm fail to handle the pixels of weak boundary. The proposed algorithm produces the desirable segmentation result for the juxta-pleural pulmonary nodules, which is shown in Fig. 8(g). The segmentation results obtained by the FCM, FCM_S, FCM_S1, FCM_S2, FLICM and GMMFCM algorithms for the GGO nodule are illustrated in Fig. 9. A lung CT slice around a GGO pulmonary nodule is shown Fig. 9(a).

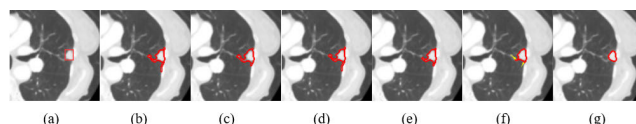


FIGURE 8. Comparison of the segmentation results using FCM, FCM_S, FCM_S1, FCM_S2, FLICM and GMMFCM algorithms for a CT image with a juxta-pleural nodule. (a) A CT image with a juxta-pleural nodule, marked with a red rectangular; (b) FCM algorithm; (c) FCM_S algorithm; (d) FCM_S1 algorithm; (e) FCM_S2 algorithm; (f) FLICM algorithm; (g) GMMFCM algorithm.

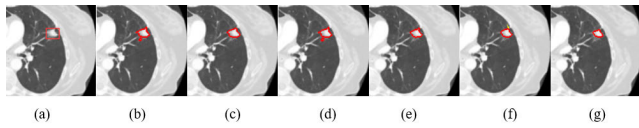


FIGURE 9. Comparison of the segmentation results using FCM, FCM_S, FCM_S1, FCM_S2, FLICM and GMMFCM algorithms for a CT image with a GGO nodule. (a) A CT image with a GGO nodule, marked with a red rectangular; (b) The segmentation result of FCM algorithm; (c) FCM_S algorithm; (d) FCM_S1 algorithm; (e) FCM_S2 algorithm; (f) FLICM algorithm; (g) GMMFCM algorithm.

Fig. The segmentation contours obtained by using FCM, FCM_S, FCM_S1, FCM_S2, FLICM algorithms are shown in 12(b)-(f), respectively. As observed in Fig. 9(b)-(e), the FCM, FCM_S, FCM_S1, FCM_S2 algorithms also produce the over-segmentation results for the GGO nodule. The segmentation result of the FLICM algorithm for a GGO pulmonary nodule are shown in Fig. 9 (f). From Fig. 9(f), we can observe that the FLICM algorithm produces slightly over-segmentation result for GGO nodule, which is indicated by yellow arrows. The negative effects of intensity inhomogeneity within the pulmonary nodule also cannot be completely resolved by the FLICM algorithm. The proposed algorithm yields the satisfactory segmentation result for the GGO pulmonary nodule, which is shown in Fig. 9(g).

2) QUALITATIVE EVALUATION ON THE GHGZMCPLA DATASET

The segmentation performance of the proposed GMMFCM algorithm is also evaluated on the GHGZMCPLA dataset. In a comparison experiment, the proposed GMMFCM algorithm is performed to segment the CT slices around various types of pulmonary nodules and compare the segmentation results with ground truths.

Fig. 10 shows the segmentation results of some examples. It is clearly seen that these examples are corrupted by intensity inhomogeneity and fuzzy boundary. For each pulmonary nodule, four images are shown in Fig. 10. They are, from left to right, the preprocessed image, the segmentation result using the proposed algorithm, the ground truth and the comparison result between the segmented nodule and ground truth. As shown in the last column of Fig. 10, the proposed algorithm produces the segmentation results closer to the ground truths. The experimental results also demonstrate that the proposed algorithm obtains the satisfactory segmentation results for different types of pulmonary nodules from the GHGZMCPLA dataset. Another comparison experiment is conducted on the GHGZMCPLA dataset to evaluate the superior performance of the proposed GMMFCM algorithm.

Fig. 11 shows the comparison results with FCM, FCM_S, FCM_S1, FCM_S2, FLICM algorithms. As can be seen from the last column of Fig. 11, the pixels belonging to nodule structures can be accurately identified. The comparisons of the algorithms can be similarly illustrated by the comparison experiments on the LIDC dataset. The qualitative results of comparison experiments on the LIDC dataset and the

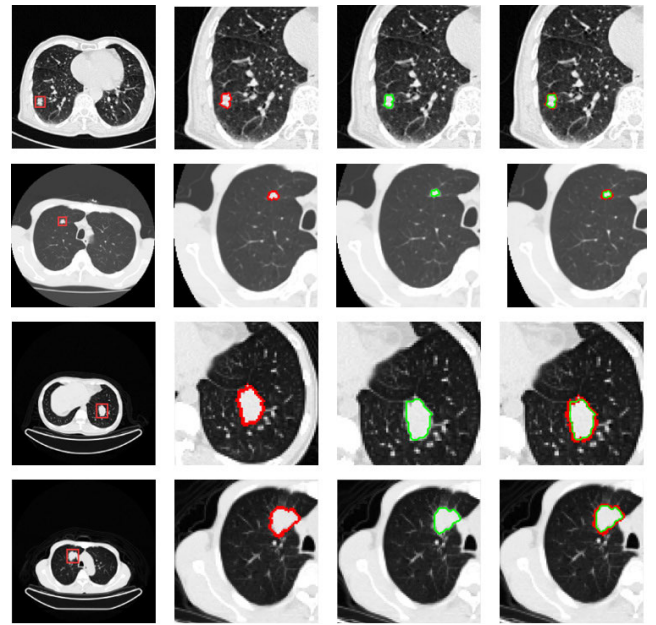


FIGURE 10. Examples of segmentation results on the GHGZMCPLA dataset using GMMFCM algorithm for different types of pulmonary nodules. The first column: The CT images after the preprocessing; The second column: the close-ups of the segmentation results; The third column: the contours of the ground truths; The fourth column: the comparisons between the segmented results and ground truths.

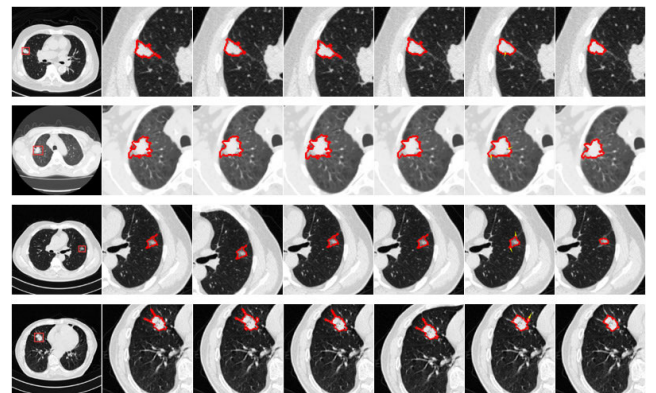


FIGURE 11. Comparisons of the segmentation results of some examples from the GHGZMCPLA dataset using FCM, FCM_S, FCM_S1, FCM_S2, FLICM algorithms. The first column: CT images with pulmonary nodules, marked with a red rectangular; The second column: FCM algorithm; The third column: FCM_S algorithm; The fourth column: FCM_S1 algorithm; The fifth column: FCM_S2 algorithm; The sixth column: FLICM algorithm; The seventh column: the proposed algorithm.

GHGZMCPLA dataset demonstrate that the proposed algorithm can achieve satisfactory results for various types of pulmonary nodules. The good segmentation results could be attributed to that the proposed GMMFCM algorithm incorporates the spatial and statistical information to reduce the boundary degradation. The proposed GMMFCM algorithm introduces the new local similar measure to define the weighting factor, which is not only influenced by the spatial information of neighboring pixels but also the posterior probability

belonging to the clusters, resulting in the improvement of segmentation results.

E. QUANTITATIVE RESULTS AND COMPARISONS

In this section, some comparison experiments are conducted on two datasets to quantitatively evaluate the segmentation accuracy and effectiveness of the proposed algorithm.

1) QUANTITATIVE EVALUATION ON THE LIDC DATASET

In a first experiment, to quantitatively evaluate the proposed GMMFCM algorithm on the LIDC dataset, we run the code of the proposed GMMFCM algorithm to segment the 100 CT slices around various types of pulmonary nodules and compare the segmentation results with ground truth. The segmentation results of only 23 cases are summarized in Table 1 due to the limited space of the paper. However, because these cases are randomly selected, the results is similar for all the considered cases. The first column and second columns show the IDs and the corresponding names. For each case, the Accuracy, Sensitivity, Specificity, Overlap score and Dice similarity coefficient (DSC) are given in columns number the three, four, five, six and seven. As observed from Table 1, an average accuracy of the proposed SRW_GMMFCM algorithm is 0.9997 ± 0.0002 . The proposed algorithm also achieves an average sensitivity of 0.9257 ± 0.0545 and an average specificity of 0.9998 ± 0.0002 , respectively. The results manifests that the proposed algorithm produces the less FPs than that of the ground truth. Fig. 12 shows the

TABLE 1. Quantitative segmentation results using the proposed GMMFCM algorithm in terms of Accuracy, Sensitivity, Specificity, Overlap, DSC.

IDs	Names	Acc	Se	Sp	Overlap	DSC
1	LIDC-0003-000044	0.9998	0.9756	0.9999	0.8889	0.9412
2	LIDC-0007-000026	0.9992	0.9504	0.9995	0.8644	0.8651
3	LIDC-0013-000096	0.9998	0.8478	1.0000	0.8387	0.9123
4	LIDC-0022-000074	0.9992	0.8673	0.9998	0.8363	0.9109
5	LIDC-0023-000088	0.9997	0.9429	0.9999	0.9086	0.9521
6	LIDC-0043-000016	0.9997	0.8858	0.9999	0.8315	0.9080
7	LIDC-0058-000073	0.9997	0.9866	0.9997	0.9095	0.9526
8	LIDC-0068-000062	0.9999	0.9434	0.9999	0.8621	0.9256
9	LIDC-0072-000196	0.9998	0.9760	0.9999	0.9314	0.9645
10	LIDC-0075-000249	0.9996	0.8723	0.9999	0.8425	0.9145
11	LIDC-0080-000215	0.9996	0.8553	1.0000	0.8497	0.9187
12	LIDC-0095-000153	0.9998	0.9744	0.9998	0.9048	0.9500
13	LIDC-0111-000080	0.9998	0.8261	1.0000	0.8143	0.8976
14	LIDC-0144-000108	0.9999	0.9615	0.9999	0.8475	0.9174
15	LIDC-0013-000063	0.9994	1.0000	0.9994	0.6866	0.8142
16	LIDC-0191-000103	0.9997	0.9900	0.9998	0.9213	0.9590
17	LIDC-0229-000105	0.9999	0.9194	1.0000	0.8906	0.9421
18	LIDC-0277-000083	0.9999	0.9375	1.0000	0.9000	0.9474
19	LIDC-0313-000028	0.9998	0.9130	1.0000	0.8972	0.9459
20	LIDC-0332-000063	0.9997	0.9477	0.9998	0.8859	0.9395
21	LIDC-0365-000074	0.9995	0.9623	0.9993	0.7402	0.8507
22	LIDC-0462-000118	0.9998	0.8209	1.0000	0.7857	0.8800
23	LIDC-0470-000010	0.9993	0.9345	0.9995	0.7772	0.8747
Mean		0.9997	0.9257	0.9998	0.8528	0.9167
Std.		0.0002	0.0545	0.0002	0.0608	0.0384

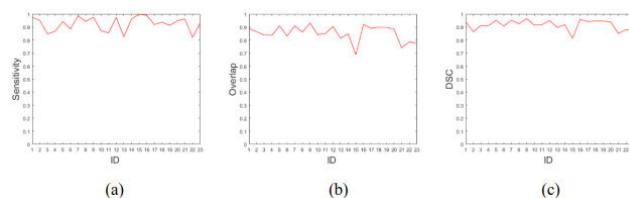


FIGURE 12. Curves of quantitative segmentation results of the proposed algorithm. (a) The curve of Sensitivity; (b) The curve of Overlap; (c) The curve of DSC.

validation results of the proposed GMMFCM algorithm on 23 cases.

Fig. 12 (a) is the sensitivity curve. A high sensitivity of the proposed algorithm is achieved by the proposed algorithm. Fig. 12 (b) and (c) are the overlap and DSC curves, respectively. From Fig. 12(b) and (c), we can observe that the proposed algorithm still remains a high overlap and a high DSC, which indicate the high similarity between the segmented nodule and ground truth.

Another comparison experiment is conducted using FCM, FCM_S, FCM_S1, FCM_S2, FLICM, FCM-type [39] and Non-FCM [40] algorithms to quantitatively evaluate the superior performance of the proposed GMMFCM algorithm on the LIDC dataset. Table 2 summarizes the average values of accuracy, sensitivity, specificity, overlap, DSC, FPR and FNR for each algorithm of the six algorithms compared above. As shown in Table 2, the performance of FCM algorithm is slightly smaller than that of FCM_S algorithm for pulmonary nodule segmentation. This is because that FCM algorithm is sensitive to the image noise and intensity inhomogeneity, while the FCM_S algorithm considers the spatial neighboring information. As shown in the fifth row of Table 2, the FCM_S1 or FCM_S2 algorithm is slightly better than FCM_S but worse than FLICM algorithm. That is because the FLICM algorithm incorporates both the local spatial information and gray level relationship. The mean accuracy of the FLICM algorithm is 0.9991 and the mean overlap is 0.8204, which indicate the segmentation performance improvement of the FLICM algorithm.

TABLE 2. Quantitative comparison of segmentation results on the LIDC dataset using FCM, FCM_S, FCM_S1, FCM_S2, FLICM, FCM-type, Non-FCM and GMMFCM algorithms in terms of Accuracy, Sensitivity, Specificity, Overlap, DSC.

Algorithms	Acc	Se	Sp	Overlap	DSC
FCM	0.9988	0.9651	0.9999	0.6399	0.7658
FCM_S	0.9986	0.9721	0.9986	0.5814	0.7125
FCM_S1	0.9989	0.9708	0.9990	0.6399	0.7658
FCM_S2	0.9989	0.9729	0.9990	0.6419	0.7740
FLICM	0.9991	0.9645	0.9993	0.7182	0.8204
FCM-type	0.9993	0.9657	0.9995	0.7468	0.8468
Non-FCM	0.9993	0.9694	0.9996	0.8145	0.8696
GMMFCM	0.9995	0.9756	0.9999	0.8581	0.8986

The mean accuracy of the proposed algorithm is 0.9995 and the mean overlap is 0.8581, which are higher than those of the FLICM algorithms for different types of pulmonary nodules. The higher overlap and DSC of the proposed algorithm indicate the higher similarity between the segmented nodule and ground truth. The sensitivity and specificity of the proposed algorithm is 0.9756 and 0.9999, which are also higher than those of the FLICM algorithm. The proposed algorithm obtains a lower FPR and a lower FNR, which also indicates the high segmentation performance. No matter which types of pulmonary nodules, the proposed algorithm still achieves a better segmentation performance than that of other five algorithms, as it uses the prior knowledge. The quantitative analysis of comparison results demonstrates that the prior knowledge can improve the segmentation performance when segmented various types of pulmonary nodules. Accordingly, as confirmed by the quantitative results, we can conclude that the incorporation of GMM posterior probability and spatial information after the enhanced images with pulmonary nodules leads to a better segmentation performance compared to the incorporation of spatial information alone. Therefore, the proposed GMMFCM algorithm can successfully segment different types of pulmonary nodules and significantly improve the performance of pulmonary nodule segmentation.

2) QUANTITATIVE EVALUATION ON THE GHGZMCPLA DATASET

To evaluate the effectiveness of the proposed GMMFCM algorithm on the GHGZMCPLA dataset, a comparison experiment is conducted using FCM, FCM_S, FCM_S1, FCM_S2, FLICM, FCM-type and Non-FCM algorithms.

Table 3 shows the average values of accuracy, sensitivity, specificity, overlap and DSC metrics between the segmented nodules and ground truth on the GHGZMCPLA dataset. As shown in Table 3, the segmentation results of the proposed algorithm with a highest Overlap and DSC are better than the others for the segmentation of various types of pulmonary nodules. In addition, the sensitivity of 0.9868 and specificity of 0.9997 are also higher than the almost others.

TABLE 3. Quantitative comparison of segmentation results on the GHGZMCPLA dataset using FCM, FCM_S, FCM_S1, FCM_S2, FLICM, FCM-type, Non-FCM and GMMFCM algorithms in terms of Accuracy, Sensitivity, Specificity, Overlap and DSC.

Algorithms	Acc	Se	Sp	Overlap	DSC
FCM	0.9987	0.9724	0.9996	0.6234	0.7432
FCM_S	0.9988	0.9835	0.9988	0.5980	0.7484
FCM_S1	0.9989	0.9817	0.9989	0.6422	0.7764
FCM_S2	0.9990	0.9587	0.9991	0.6739	0.7759
FLICM	0.9994	0.9737	0.9995	0.7500	0.8571
FCM-type	0.9994	0.9786	0.9996	0.7687	0.8764
Non-FCM	0.9995	0.9794	0.9996	0.8045	0.9045
GMMFCM	0.9996	0.9868	0.9997	0.8743	0.9330

The experimental results manifest the robustness and effectiveness of the proposed algorithm on GHGZMCPLA dataset for pulmonary nodule segmentation.

F. COMPARISON WITH THE STATE-OF-THE-ART ALGORITHMS

To justify the accuracy and effectiveness of the proposed algorithm for pulmonary nodule segmentation, we compare the proposed algorithm with the state-of-the-art segmentation algorithms of pulmonary nodules. It is difficult to make a comparison with previously published literatures, as some of them usually use the private datasets or the subsets of public datasets, and sometimes, the published literature do not report the cases that are selected. However, it is very important to stress a fair comparison among the algorithms. Therefore, we attempt to compare the published performance results that use the LIDC dataset, which helps to mitigate one of the variability factors. Since some metrics used in different algorithms are inconsistent in the published literatures, the mainly reported overlap scores of these state-of-the-art algorithms are listed in Table 4 for a completely fair comparison. We believe that the performance comparison results on the GHGZMCPLA dataset are sufficiently similar to those of the LIDC dataset.

TABLE 4. Comparison of segmentation performance on the LIDC dataset with the state-of-the-art algorithms. Overlap scores are relative to the 50% consensus truth.

Authors	Overlap
W. J. Kostis <i>et al.</i> (2003) [50]	0.57 ± 0.20
K. Okada <i>et al.</i> (2005) [6]	0.52 ± 0.25
B. van Ginneken <i>et al.</i> (2006) [51]	0.66 ± 0.18
J. M. Kuhnigk <i>et al.</i> (2006) [52]	0.67 ± 0.22
J. Wang <i>et al.</i> (2007) [53]	0.64 ± 0.18
T. Messay <i>et al.</i> (2010) [54]	0.63 ± 0.16
T. Kubota <i>et al.</i> (2011) [4]	0.69 ± 0.18
T. Messay <i>et al.</i> (2015) [55]	0.77 ± 0.09
Our algorithm	0.86 ± 0.06

Here, each row represents a published algorithm for pulmonary nodule segmentation and lists the reported overlap score. The first through eighth rows shown the reported overlap score of eight algorithms. They are, from up to down, the algorithms reported by Kostis *et al.* [50], Okada *et al.* [6], van Ginneken [51], Kuhnigk *et al.* [52], Wang *et al.* [53], Messay *et al.* [54], Kubota *et al.* [4] and Messay *et al.* [55]. Their algorithms will be called W. J. Kostis's algorithm, K. Okada's algorithm, B. van Ginneken's algorithm, J. M. Kuhnigk's algorithm, J. Wang's algorithm, T. Messay2010's algorithm, T. Kubota's algorithm and T. Messay2015's algorithm for the convenience of description. The average overlap of W. J. Kostis's algorithm and K. Okada's algorithm were 0.57 ± 0.20 and 0.52 ± 0.25 , respectively. W. J. Kostis's algorithm obtained a lowest overlap among the other algorithms. This was

because that W. J. Kostis's algorithm used an ellipsoid shape to model the pulmonary nodules. However, the pulmonary nodule boundaries were not strictly ellipsoid shape, which might miss some parts of real pulmonary nodule boundaries. The average overlap of B. van Ginneken's algorithm was 0.66 ± 0.18 , which was higher than that of W. J. Kostis's algorithm and K. Okada's algorithm. This was because that B. van Ginneken's algorithm used the leave-one-out regime. However, it required the manual classification of pulmonary nodule into solid and non-solid nodules. The average overlap of J. M. Kuhnigk's algorithm and J. Wang's algorithm were 0.67 ± 0.22 and 0.64 ± 0.18 , respectively. J. M. Kuhnigk's algorithm obtained a relative higher overlap than the above other algorithms. This was because that J. M. Kuhnigk's algorithm used an erosion operation to remove the adjacent vessel structures. However, it obtained a lower higher overlap than that of T. Messay2015's algorithm. Since it was difficult to determine the size of structure element, too large or too small size could have the great impact for segmentation accuracy. The erosion operation remove may remove a significant portion of the pulmonary nodule, because the narrowest part of nodule and the adjacent vessel have the same size. T. Kubota's algorithm obtained a higher overlap of 0.69 ± 0.18 than that of T. Messay2010's algorithm. The authors took full advantage of additional control points to improve the segmentation performance. However, this algorithm was inadequate to segment the invasive juxta-pleural nodules. T. Messay2015's algorithm reported the average overlap of 0.77 ± 0.09 , which was a relative higher overlap than that of the T. Kubota's algorithm and was comparable to our algorithm. The competition-diffusion (CD) based figure-ground separation could effectively remove the partial volume effects, resulting in the better segmentation performance.

As observed from the last row of Table 4, the proposed algorithm achieves a competitive segmentation performance in comparison with other algorithms. Due to the similar intensity between the nodule and adjacent structures, the segmentation of juxta-vascularized nodules and juxta-pleural nodules is usually not so accurate by using these algorithms. It is worth noting that the proposed algorithm achieves the average overlap of 0.86 ± 0.06 , which is the highest overlap scores among these state-of-the-art segmentation algorithms. The good segmentation performance of the proposed algorithm attributes to the incorporate of GMM prior knowledge and spatial information.

To evaluate the computational efficiency, the mean executive time is summarized in Table 5, which is measured in seconds on LIDC datasets. 23 cases with the nodules are selected in our experiment, which provide with different shapes, sizes and texture information. We run our code in 23 cases. The GMMFCM algorithm and Kostis's algorithm [50] obtain the mean executive time of 3.14 and 4.49, respectively. The executive time of GMMFCM algorithm is notably lower than other state-of-the-art algorithms. The result demonstrates that the proposed method is much faster than other state-of-the-art algorithms.

TABLE 5. Comparison of executive time on the LIDC dataset with the state-of-the-art algorithms.

Authors	Times(s)
W. J. Kostis et al. (2003) [50]	4.49
K. Okada et al. (2005) [6]	4.37
B. van Ginneken et al. (2006) [51]	4.98
J. M. Kuhnigk et al. (2006) [52]	5.63
J. Wang et al. (2007) [53]	4.76
T. Messay et al. (2010) [54]	3.69
T. Kubota et al. (2011) [4]	4.28
T. Messay et al. (2015) [55]	3.47
Our algorithm	3.14

V. DISCUSSIONS

The proposed GMMFCM algorithm can successfully segment various types of pulmonary nodules. The proposed superpixel-based random walker could reduce the computational complexity and the risk of assigning error labels to the corresponding pixels. The proposed multiscale dot enhancement filter could generate more reliable nodule seeds, and the proposed scheme of background seeds generation could generate more accurate background. The nodule and background GMM models were built by using the generated seeds. Then, the GMM models were used to define a local similarity measure. The new energy function was defined based on the local similarity measure to improve the segmentation performance. The comparison visual segmentation results of the proposed algorithm compared with the ground truths on two datasets were shown Fig. 6 and Fig. 10. For some complex cases, such as juxta-pleural, juxta-vascularized and GGO nodules, the comparison results with FCM, FCM_S, FCM_S1, FCM_S2 and LIFCM algorithms on the LIDC dataset were shown in Fig. 7, Fig. 8 and Fig. 9. The comparison results on the GHGZMCPA dataset were shown in Fig. 11. The segmentation results were shown by the red contours and the ground truths were shown by the green contours.

The quantitative comparison was performed in terms of seven evaluation criteria, including accuracy (see Eqn. 13), sensitivity (see Eqn. 14), specificity (see Eqn. 15), FPR (see Eqn. 16) and FNR (see Eqn. 17), overlap score (see Eqn. 18) and DSC (see Eqn. 19). For comparison purpose, 100 images from the LIDC dataset were selected and each image was rescaled to 256×256 pixels to reduce the time cost. Considering the limited space of the paper, the five evaluation criteria of only 23 cases were shown in Table 1. Based on the numerical results shown in Table 1, it could be seen that the proposed algorithm was capable of segmenting various types of pulmonary nodules and obtained the desired segmentation results. The average overlap and DSC were 0.8528 and 0.9167, respectively.

In particular, the overlap and DSC were compared in Fig. 12(b) and (c), respectively. The sensitivity of pulmonary nodule segmentation were shown in Fig. 12(a).

For all cases, the quantitative segmentation results of the proposed algorithm were compared with the results provided by the FCM, FCM_S, FCM_S1, FCM_S2 and LIFCM algorithms, as well as some other state-of-the-art algorithms of pulmonary nodule segmentation. In particular, the LIFCM algorithm was comparable to the proposed algorithm. The comparison results of proposed algorithm with the conventional FCM, FCM_S, FCM_S1, FCM_S2 and LIFCM, FCM-type and Non-FCM algorithms on the LIDC dataset and the GHGZMCPLA dataset were shown in Table 2 and Table 3, respectively. From Table 2 it can be clearly seen that the proposed algorithm provided the high overlap about 0.8581 and DSC about 0.8986, which indicated a high similarity between the segmentation results and ground truth. The similar results could be seen from Table 3, which compared the results on the GHGZMCPLA dataset. The segmentation performance of the proposed algorithm was improved by using the new local similar measure. The quantitative segmentation results of the proposed algorithm were compared with the results provided by the state-of-the-art algorithms of pulmonary nodule segmentation. Based on the numerical results shown in Table 4 it can be also seen that the proposed algorithm achieved the comparable segmentation results.

VI. CONCLUSION AND FUTURE WORK

In this paper, the GMMFCM algorithm is proposed for segmentation of the pulmonary nodules. We performed a detailed segmentation performance comparison with FCM, FCM_S, FCM_S1, FCM_S2, FLICM, FCM-type and Non-FCM algorithms. The experiments have been conducted on the LIDC dataset and the GHGZMCPLA dataset to test the performance of the GMMFCM algorithm. The results show that the performance of the proposed GMMFCM algorithm is promising for segmentation of the pulmonary nodules, and is more robust than the other algorithms.

Several other factors also play a role in the performance improvement of the proposed algorithm. Superpixel-based random walker algorithm is employed for pulmonary parenchyma segmentation, and a new multiscale dot enhancement filter is defined for nodule seed generation. The local similarity measure is defined by using the GMM posterior probability and spatial information. In particular, the overlap and sensitivity of the proposed algorithm are the higher compared with some other algorithms.

In future work, we will investigate the techniques of data sampling so that the proposed algorithm can be extended to large-scale segmentation problems. Further study on this topic will also include many applications of the GMMFCM algorithm in other problems.

REFERENCES

- [1] A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, and M. J. Thun, "Cancer statistics," *CA, Cancer J. Clinicians*, vol. 67, no. 1, 2017.
- [2] S. G. Armato, M. L. Giger, and H. Macmahon, "Automated detection of lung nodules in CT scans: Preliminary results," *Med. Phys.*, vol. 28, no. 8, pp. 1552–1561, Aug. 2001.
- [3] J. Dehmeshki, H. Amin, M. Valdivieso, and X. Ye, "Segmentation of pulmonary nodules in thoracic CT scans: A region growing approach," *IEEE Trans. Med. Imag.*, vol. 27, no. 4, pp. 467–480, Apr. 2008.
- [4] T. Kubota, A. K. Jerebko, M. Dewan, M. Salganicoff, and A. Krishnan, "Segmentation of pulmonary nodules of various densities with morphological approaches and convexity models," *Med. Image Anal.*, vol. 15, no. 1, pp. 133–154, Feb. 2011.
- [5] H. H. Jo, H. Hong, J. M. Goo, C. M. Park, and M. H. Lee, "Automatic segmentation of pulmonary nodule in CT images using intensity and morphology-based hybrid algorithm," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 5, pp. S354–S374, Jun. 2010.
- [6] K. Okada, D. Comaniciu, and A. Krishnan, "Robust anisotropic Gaussian fitting for volumetric characterization of Pulmonary nodules in multislice CT," *IEEE Trans. Med. Imag.*, vol. 24, no. 3, pp. 409–423, Mar. 2005.
- [7] L. Gonçalves, J. Novo, and A. Campilho, "Hessian based approaches for 3D lung nodule segmentation," *Expert Syst. Appl.*, vol. 61, pp. 1–15, Nov. 2016.
- [8] C. Zhou, H.-P. Chan, B. Sahiner, L. M. Hadjiiski, A. Chughtai, S. Patel, J. Wei, J. Ge, P. N. Cascade, and E. A. Kazerooni, "Automatic multiscale enhancement and segmentation of pulmonary vessels in CT pulmonary angiography images for CAD applications," *Med. Phys.*, vol. 34, no. 12, pp. 4567–4577, Nov. 2007.
- [9] K. Chen, B. Li, L.-F. Tian, W.-B. Zhu, and Y.-H. Bao, "Vessel attachment nodule segmentation using integrated active contour model based on fuzzy speed function and shape–intensity joint Bhattacharya distance," *Signal Process.*, vol. 103, pp. 273–284, Oct. 2014.
- [10] B. Li, K. Chen, G. Peng, Y. Guo, L. Tian, S. Ou, and L. Wang, "Segmentation of ground glass opacity pulmonary nodules using an integrated active contour model with wavelet energy-based adaptive local energy and posterior probability-based speed function," *Mater. Express*, vol. 6, no. 4, pp. 317–327, Aug. 2016.
- [11] B. Li, Q. Chen, G. Peng, Y. Guo, K. Chen, L. Tian, S. Ou, and L. Wang, "Segmentation of pulmonary nodules using adaptive local region energy with probability density function-based similarity distance and multi-features clustering," *Biomed. Eng. Online*, vol. 15, no. 1, May 2016.
- [12] A. A. Farag, H. E. A. El Munim, J. H. Graham, and A. A. Farag, "A novel approach for lung nodules segmentation in chest CT using level sets," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5202–5213, Dec. 2013.
- [13] J. S. Schildkraut, N. Prosser, A. Savakis, J. Gomez, D. Nazareth, A. K. Singh, and H. K. Malhotra, "Level-set segmentation of pulmonary nodules in megavolt electronic portal images using a CT prior," *Med. Phys.*, vol. 37, no. 11, pp. 5703–5710, Oct. 2010.
- [14] H. Liu, C.-M. Zhang, Z.-Y. Su, K. Wang, and K. Deng, "Research on a pulmonary nodule segmentation method combining fast self-adaptive FCM and classification," *Comput. Math. Methods Med.*, vol. 2015, Apr. 2015, Art. no. 185726.
- [15] M. Kakar and D. R. Olsen, "Automatic segmentation and recognition of lungs and lesion from CT scans of thorax," *Computerized Med. Imag. Graph.*, vol. 33, no. 1, pp. 72–82, Jan. 2009.
- [16] S. Diciotti, S. Lombardo, M. Falchini, G. Picozzi, and M. Mascalchi, "Automated segmentation refinement of small lung nodules in CT scans by local shape analysis," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 12, pp. 3418–3428, Dec. 2011.
- [17] M. Keshani, Z. Azimifar, F. Tajeripour, and R. Boostani, "Lung nodule segmentation and recognition using SVM classifier and active contour modeling: A complete intelligent system," *Comput. Biol. Med.*, vol. 43, no. 4, pp. 287–300, May 2013.
- [18] S. M. B. Netto, A. Corrêa Silva, R. A. Nunes, and M. Gattass, "Automatic segmentation of lung nodules with growing neural gas and support vector machine," *Comput. Biol. Med.*, vol. 42, no. 11, pp. 1110–1121, Nov. 2012.
- [19] X. Chang, Q. Wang, Y. Liu, and Y. Wang, "Sparse regularization in fuzzy c-means for high-dimensional data clustering," *IEEE Trans. Cybern.*, vol. 47, no. 9, pp. 2616–2627, Sep. 2017.
- [20] D. D. Nguyen, L. T. Ngo, L. T. Pham, and W. Pedrycz, "Towards hybrid clustering approach to data classification: Multiple kernels based interval-valued Fuzzy C-Means algorithms," *Fuzzy Sets Syst.*, vol. 279, pp. 17–39, Nov. 2015.
- [21] K.-P. Lin, "A novel evolutionary kernel intuitionistic fuzzy C-means clustering algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 5, pp. 1074–1087, Oct. 2014.
- [22] F. Zhao, Z. Zeng, H. Q. Liu, and J. L. Fan, "A Kriging-assisted reference vector guided multi-objective evolutionary fuzzy clustering algorithm for image segmentation," *IEEE Access*, vol. 7, pp. 21465–21481, 2019.

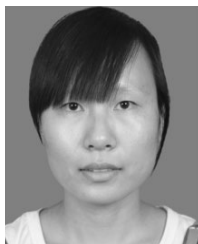
- [23] H. Huang, F. Meng, S. Zhou, F. Jiang, and G. Manogaran, "Brain image segmentation based on FCM clustering algorithm and rough set," *IEEE Access*, vol. 7, pp. 12386–12396, 2019.
- [24] A. A. Othman, H. R. Tizhoosh, and F. Khalvati, "EFIS—Evolving fuzzy image segmentation," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 1, pp. 72–82, Feb. 2014.
- [25] S. Balla-Arabe, X. Gao, and B. Wang, "A fast and robust level set method for image segmentation using fuzzy clustering and lattice Boltzmann method," *IEEE Trans. Cybern.*, vol. 43, no. 3, pp. 910–920, Jun. 2013.
- [26] L. Chen, C. L. P. Chen, and M. Lu, "A multiple-kernel fuzzy C-means algorithm for image segmentation," *IEEE Trans. Syst., Man, B, Cybern.*, vol. 41, no. 5, pp. 1263–1274, Oct. 2011.
- [27] Y. Yi, W. Zhou, and C. Bi, "Inner product regularized nonnegative self representation for image classification and clustering," *IEEE Access*, vol. 43, pp. 14165–14176, 2017.
- [28] H. Van Assen, M. Danilouchkine, M. Dirksen, J. Reiber, and B. Lelieveldt, "A 3-D active shape model driven by fuzzy inference: Application to cardiac CT and MR," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 5, pp. 595–605, Sep. 2008.
- [29] X. Yang and B. Fei, "A multiscale and multiblock fuzzy C-means classification method for brain MR images," *Med. Phys.*, vol. 38, pp. 2879–2891, May 2011.
- [30] Y.-C. Chang, Y.-H. Huang, C.-S. Huang, P.-K. Chang, J.-H. Chen, and R.-F. Chang, "Classification of breast mass lesions using model-based analysis of the characteristic kinetic curve derived from fuzzy c-means clustering," *Magn. Reson. Imag.*, vol. 30, no. 3, pp. 312–322, Apr. 2012.
- [31] A. Arshad, S. Riaz, and L. Jiao, "Semi-supervised deep fuzzy C-mean clustering for imbalanced multi-class classification," *IEEE Access*, vol. 7, pp. 28100–28112, 2019.
- [32] I. Gatos, S. Tsantis, M. Karamesini, S. Spiliopoulos, D. Karnabatidis, and J. D. Hazle, "Focal liver lesions segmentation and classification in non-enhanced T2-weighted MRI," *Med. Phys.*, vol. 44, no. 7, pp. 3695–3705, Jul. 2017.
- [33] W. Cai, S. Chen, and D. Zhang, "Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation," *Pattern Recognit.*, vol. 40, no. 3, pp. 825–838, Mar. 2007.
- [34] S. Krinidis and V. Chatzis, "A robust fuzzy local information C-means clustering algorithm," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1328–1337, May 2010.
- [35] N. Li, H. Huo, Y.-M. Zhao, X. Chen, and T. Fang, "A spatial clustering method with edge weighting for image segmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1124–1128, Sep. 2013.
- [36] M. Gong, Y. Liang, J. Shi, W. Ma, and J. Ma, "Fuzzy C-means clustering with local information and kernel metric for image segmentation," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 573–584, Feb. 2013.
- [37] Z. Ji, J. Liu, G. Cao, Q. Sun, and Q. Chen, "Robust spatially constrained fuzzy c-means algorithm for brain MR image segmentation," *Pattern Recognit.*, vol. 47, no. 7, pp. 2454–2466, Jul. 2014.
- [38] H. Zhang, Q. Wang, W. Shi, and M. Hao, "A novel adaptive fuzzy local information C-means clustering algorithm for remotely sensed imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5057–5068, Sep. 2017.
- [39] F. Zhao, Y. Chen, H. Liu, and J. Fan, "Alternate PSO-based adaptive interval type-2 intuitionistic fuzzy C-means clustering algorithm for color image segmentation," *IEEE Access*, vol. 7, pp. 64028–64039, 2019.
- [40] C. Singh and A. Bala, "A DCT-based local and non-local fuzzy C-means algorithm for segmentation of brain magnetic resonance images," *Appl. Soft Comput.*, vol. 68, pp. 447–457, Jul. 2018.
- [41] S. Zhou, Y. Cheng, and S. Tamura, "Automated lung segmentation and smoothing techniques for inclusion of juxtapleural nodules and pulmonary vessels on chest CT images," *Biomed. Signal Process. Control*, vol. 13, pp. 62–70, Sep. 2014.
- [42] A. Mansoor, U. Bagci, Z. Xu, B. Foster, K. N. Olivier, J. M. Elinoff, A. F. Suffredini, J. K. Udupa, and D. J. Mollura, "A generic approach to pathological lung segmentation," *IEEE Trans. Med. Imag.*, vol. 33, no. 12, pp. 2293–2310, Dec. 2014.
- [43] A. Eslami, A. Karamalis, A. Katouzian, and N. Navab, "Segmentation by retrieval with guided random walks: Application to left ventricle segmentation in MRI," *Med. Image Anal.*, vol. 17, no. 2, pp. 236–253, Feb. 2013.
- [44] H. Mi, C. Petitjean, P. Vera, and S. Ruan, "Joint tumor growth prediction and tumor segmentation on therapeutic follow-up PET images," *Med. Image Anal.*, vol. 23, no. 1, pp. 84–91, Jul. 2015.
- [45] W. Ju, D. Xiang, B. Zhang, L. Wang, I. Kopriva, and X. Chen, "Random walk and graph cut for co-segmentation of lung tumor on PET-CT images," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5854–5867, Dec. 2015.
- [46] S. H. Park, S. Lee, I. D. Yun, and S. U. Lee, "Structured patch model for a unified automatic and interactive segmentation framework," *Med. Image Anal.*, vol. 24, no. 1, pp. 297–312, Aug. 2015.
- [47] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
- [48] Q. Li, S. Sone, and K. Doi, "Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans," *Med. Phys.*, vol. 30, no. 8, pp. 2040–2051, Jul. 2003.
- [49] X. Ye, X. Lin, J. Dehmeshki, G. Slabaugh, and G. Beddoe, "Shape-based computer-aided detection of lung nodules in thoracic CT images," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 7, pp. 1810–1820, Jul. 2009.
- [50] W. Kostis, A. Reeves, D. Yankelevitz, and C. Henschke, "Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical ct images," *IEEE Trans. Med. Imag.*, vol. 22, no. 10, pp. 1259–1274, Oct. 2003.
- [51] B. van Ginneken, "Supervised probabilistic segmentation of pulmonary nodules in CT scans," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 4191, 2016, pp. 912–919.
- [52] J.-M. Kuhnigk, V. Dicken, L. Bornemann, A. Bakai, D. Wormanns, S. Krass, and H.-O. Peitgen, "Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic CT scans," *IEEE Trans. Med. Imag.*, vol. 25, no. 4, pp. 417–434, Apr. 2006.
- [53] J. Wang, R. Engelmann, and Q. Li, "Segmentation of pulmonary nodules in three-dimensional CT images by use of a spiral-scanning technique," *Med. Phys.*, vol. 34, no. 12, pp. 4678–4689, Nov. 2007.
- [54] T. Messay, R. C. Hardie, and S. K. Rogers, "A new computationally efficient CAD system for pulmonary nodule detection in CT imagery," *Med. Image Anal.*, vol. 14, no. 3, pp. 390–406, Jun. 2010.
- [55] T. Messay, R. C. Hardie, and T. R. Tuinstra, "Segmentation of pulmonary nodules in computed tomography using a regression neural network approach and its application to the lung image database consortium and image database resource initiative dataset," *Med. Image Anal.*, vol. 22, no. 1, pp. 48–62, May 2015.



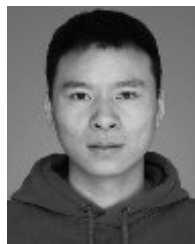
XIANGXIA LI received the Ph.D. degree from the South China University of Technology, Guangzhou, China. She is currently a Lecturer with the Information Engineering, Guangdong University of Finance and Economics, Guangzhou, China. Her current research interests include medical image segmentation, medical image processing, and pattern recognition.



BIN LI received the B.S. and Ph.D. degrees from the School of Automation Science and Engineering, South China University of Technology (SCUT), Guangzhou, China, in 2002 and 2007, respectively. He is currently an Associate Professor of automation science and engineering with SCUT. His current research interests include information visualization, medical image processing, and pattern recognition.



FANG LIU received the Ph.D. degree from the South China University of Technology, Guangzhou, China. She is currently a Lecturer with the School of Internet Finance and Information Engineering, Guangdong University of Finance and Economics. Her research interests include transfer learning, privileged information, and action recognition.



FENG ZHOU received the B.S. degree in information computing science from Minnan Normal University, China, in 2010, and the Ph.D. degree in computational mathematics from Sun Yat-sen University, China, in 2015. As a Visiting Student, he studied at the School of Mathematics, Georgia Institute of Technology, from September 2013 to September 2014. He is currently an Assistant Professor with the School of Information, Guangdong University of Finance and Economics. His research interests include signal analysis, machine learning, and ensemble learning.

• • •



HUA YIN received the Ph.D. degree from the Wuhan University, Wuhan, China. She is currently a Lecturer with the Information Engineering, Guangdong University of Finance and Economics, Guangzhou, China. Her current research interests include image segmentation, image processing, and pattern recognition.