Advances in Engineering, volume 126
5th International Conference on Machinery, Materials and Computing Technology (ICMMCT 2017)

Research on Modified Fuzzy C-means Algorithm in Lung Nodules Computer-aided Diagnosis (CAD) System

Qing Li[1, a], Hui Liu[2,3, b]

[1] Department of Mechanical Engineering, Shandong Labor Vocational and Technical College, Jinan, 250022, China

[2] Shandong Provincial Key Laboratory of Digital Media Technology, Jinan, 250014, China

[3] School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan, 250014, China

[a]email: liqing1995@163.com, [b]email:liuh_lh@126.com

Keywords: CAD; Fuzzy C-means Algorithm; Punishment Factor; Neighborhood Space Window; Gray Scale

Abstract. It is important for the early diagnosis and treatment of lung cancer in the Computer-aided Diagnosis/Detection (CAD) system, and accurate segmentation of pulmonary nodules from tomographic images is the basic and active research problem for the benign or malign diagnosis. For this reason, this work seeks to develop automatic detection and classification method of lung nodules. First, the algorithm separates lung parenchyma from the anatomical structures based on maximum between-cluster variance, image dilation and erosion. Secondly, a modified robust fuzzy c-means clustering(rFCM) segmentation algorithm is proposed, this method improves the objective function by adding a punishment factor, for eliminating the influence from noise and non-uniform gray problem. Experimental results have shown that the proposed method can achieve more accurate segmentation and perform better than other traditional algorithms in classification and recognition, Furthermore, the segmentation results on brain images also get a satisfied performance.

Introduction

American National Cancer Institute statistics that almost 22.7% cancer deaths is caused by lung cancer, and pulmonary nodules are potential manifestation of lung cancer [8]. So, early diagnosis is one of the key issues that can reduce the mortality of lung cancer [5]. Therefore, research on the lung cancer computer-aided diagnosis techniques has become more and more important. For promoting the development of CAD technology in lung cancer [2][3], the National Cancer Institute (NCI) of The United States established a lung CT image database---Lung Imaging Database Consortium (LIDC), which contains 1012 chest cases and 1356 nodules at the time of this study [4][13]. Each nodule case contains the painting boundary coordinates and vision features from four radiologists. And the malignancy ratings were set by the visual decision of the four radiologists from one to five. The bigger value represents more malignant suspicious.

The main working process of the CAD system is as follows: (1) the segmentation of lung nodules; (2) features extraction; (3) classification and recognition. The key technique includes image processing and machine learning. In this paper, the main method of lung nodules segmentation is based on FCM algorithm, which has been widely used in machine learning area.

The FCM algorithm is accord with human cognitive characteristics, which are simple to describe and easy to implement. But when it is applied to the field of image segmentation, the traditional FCM algorithm[7] is sensitive to noise, low in neighborhood information correlation, and easy to fall into local convergence. To solve these problems, many scholars have studied on it. Ahmed [1] etc. proposed a FCM algorithm of offset correction (FCM_S), which introduces spatial information by modifying the objective function of traditional FCM algorithm. The algorithm achieves satisfactory results in MRI image segmentation processing. Szilagyi [12] etc. proposed an enhanced FCM (EnFCM) algorithm to speed up the clustering process. It uses a linearly weighted sum image formed from both original image and each pixel's local neighborhood average gray level. In addition, clustering is performed on the basis of the gray level histogram instead of pixels of the summed image. Hence, the computational time of EnFCM algorithm is reduced greatly. A novel robust fuzzy local information C-means clustering (FLICM) [11] algorithm with local spatial and gray level information incorporated is free of the empirically adjusted parameter and enhances the clustering performance.

Based on the analysis of above problems, this paper presents a modified robust fuzzy c-means clustering(rFCM) based on FCM. The algorithm makes use of the eight neighborhood pixels to analyze the center, and takes into account the gray level information of the pixels in the window. This method reduces the sensitivity of the algorithm to noise, especially the medical image with noise, which has better segmentation effects.

Segmentation of lung nodules based on rFCM algorithm

Traditional FCM algorithm

The traditional FCM algorithm is an important branch of unsupervised pattern classification in statistical pattern recognition. We set up a limited set of $X = \{x_1, x_2, \cdots x_n\}$ consisting of n samples. The clustering problem is to divide the set X into c subsets. The samples in the same subset are as similar as possible, but the samples in different subsets as different as possible, c is the number of clusters.

$$J_{FCM} = \sum_{i=1}^{N}\sum_{j=1}^{c} u_{ij}^{m} d^2(x_i, v_j) \tag{1}$$

Where $c \in [2, N)$ is the number of clusters, u_{ij} is the fuzzy membership of the x_i sample with respect to the v_i cluster center; $m \in [2, \infty)$ is the weighting exponent; $d^2(x_i, v_j)$ denotes Euclidean distance between the x_i sample and the v_j cluster center, define as: $d^2(x_i, v_j) = \left\| x_i - v_j \right\|^2$. In order to obtain the optimal partition of the sample set X , the minimize constraint of the objective function (1) is:

$$u_{ij} \in [0,1], \sum_{i=1}^{c} u_{ij} = 1 \tag{2}$$

By using the standard technique of Lagrange Multipliers to minimize the objective function under the constraint of (2), we obtain the formula of the fuzzy membership and the clustering center:

$$v_j = \frac{\sum_{i=1}^{N} (u_{ij})^m x_i}{\sum_{i=1}^{N} (u_{ij})^m} \qquad u_{ij} = \frac{1}{\sum_{k=1}^{c} (\frac{d_{ij}}{d_{jk}})^{\frac{2}{(m-1)}}} \tag{3}$$

Robust Fuzzy Clustering Algorithm (rFCM)

In this section, we propose rFCM algorithm, which is applied in the segmentation of lung nodules image. The rFCM algorithm combines FCM_S algorithm's reference of neighborhood information and the FLICM's insensitivity to noise. Of course, there are some excellent for noise algorithms, such as contourlet [10], linear filter [6]. We take into account the characteristics of the objective function FCM by choosing the reference of neighborhood information, and improve the objective function for overcoming the problem of image excessive segmentation. The improved objective function combines the spatial characteristics and the high gray correlation of the neighborhood pixels, it predicts the gray correlation specific to the neighborhood information and applies the different clustering fuzzy factor according to correlation anticipation.

x_r	x_r	x_r
x_r	x_i	x_r
x_r	x_r	x_r

Fig.1 The center pixel's 3*3 neighborhood space window N_r

Fig.1 aggregates the gray value of the center pixel x_i 's neighborhood space window pixels, denoted by $gray(N_i)$. This method maps the image gray-scale (0~255) into new gray-scale according to the number of cluster, then judges the level of center pixel x_i in the new gray-scale, and counts the number of pixels in the neighborhood space which are in the same level with x_i, denoted by θ. It's necessary to judge whether the number θ is in the defined reference range (we define the number that satisfy in the same level with x_i as β). If $\theta < \beta$, we can see that the pixel x_i doesn't cause abrupt change of the neighborhood space, then it can be applied the fuzzy factor N_{ij} , otherwise, the pixel x_i may be noise, which given by the fuzzy factor P_{ij} .

In the conventional FCM algorithm, the weight factor of the membership m has little effect on the results. When the value of m increases, the running time increases. Considering the time efficiency of the rFCM algorithm, the weight factor m is set to 2, which is conform to the previous experience summary. The rFCM algorithm's objective function is modified as:

$$J_{rFCM} = \sum_{i=1}^{N}\sum_{j=1}^{c}[u_{ij}^2 d^2(x_i, v_j) + b_i N_{ij} + \overline{b}_i P_{ij}] \tag{5}$$

Where x_i is the gray value of pixel in the image; v_j is the center of the jth cluster; u_{ik} is the fuzzy membership of the pixel x_i with respect to the jth cluster center; N_R is the number of samples in neighborhood space; $d^2(x_i, v_j)$ denotes Euclidean distance between the pixel x_i and the jth cluster center, $b_i(i = 1, 2, \cdots, N)$ is a Boolean indicator vector with 0-1, the value of the pixels meeting the gray level that are defined as 1 and 0 otherwise; N_{ij} is the fuzzy factor that defined to meet the judgment of the pixel number in the same gray scale.

$$N_{ij} = u_{ij}^2 \frac{\alpha}{N_R} \sum_{\substack{r \in N_i \\ \theta < \beta}} \left\| x_r - v_j \right\|^2 \tag{6}$$

Where α is a parameter controlling the affect intensity of the neighborhood information; N_R is the number of samples in neighborhood space; x_r is the gray value of the neighborhood pixel. In the opposite case, rFCM uses fuzzy factor P_{ij} :

$$P_{ij} = u_{ij}^2 \sum_{\substack{x_r \in N_i \\ \theta \geq \beta}} (1 - u_{rj})^2 \left\| x_r - v_j \right\|^2 \tag{7}$$

In formula (7), x_r is the gray value of the neighborhood pixels. Each pixel x_i should be subject to the constraint of formula (2). P_{ij} is a new bound term that constrains weather the pixel x_i completely belong to the jth clustering center or not. P_{ij} avoids the ambiguity of the membership, makes the attribution of pixels clearer. P_{ij} not only takes the neighborhood information into consideration, but also constrains the value of the fuzzy membership, so we define the number that satisfying the same level with x_i as β. In formula (6) and (7), the parameter β is defined as the number that satisfying in the same level with x_i. Experiments verify that if the β is too big, the reference value of neighborhood pixels will be reduced, if the β is too small, the noise will be introduced. So the values of β should be selected according to actual condition.

Experimental results

This section aims to verify the anti-noise performance of the improved algorithm in lung CT image segmentation. Thirty lung CT images are screened out from LIDC to carry out image segmentation experiments. The segmentation results of the new algorithm are compared with the existing four typical algorithms (FCM, FCM_S, EnFCM, FLICM), as shown in Fig.2.

Fig.2 (a) is a special case that contains noise and in low definition, when compared with the traditional algorithms, we find that FCM, FCM_S, EnFCM all have obvious jaggies in the edges, and do not make effective segmentation with some small texture. Although FLICM algorithm exhibits a good segmentation result in noise images, but it obviously has over-segmentation. Our algorithm makes an accurate segmentation with the image, especially the fine reserve of the texture and the reduction of the noise sensitivity. According to the comprehensive analysis, we can come to the conclusion that the rFCM algorithm can get satisfactory results with the noise and blurred image. Furthermore, we compare the segmentation results on brain CT image, which also verify the performance of our method.

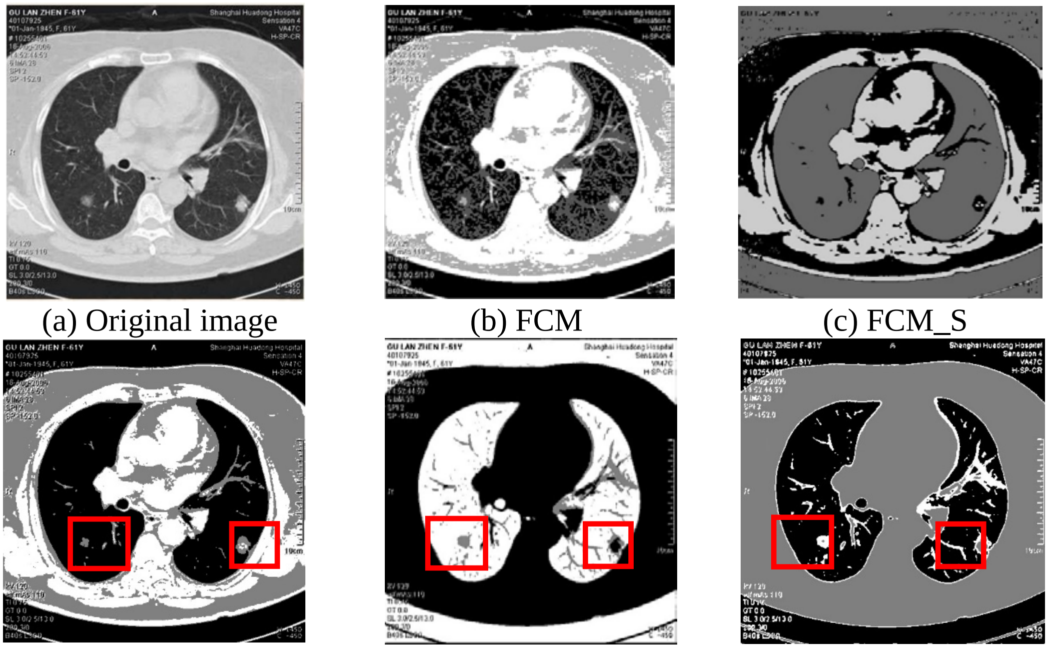
Table 1 shows the average segment precision of four kinds of typical algorithms and rFCM algorithm on the lung CT image set.

Table 1 Values of the average segment precision for the compared segmentation algorithms

Algorithm	FCM	FCM_S	EnFCM	FLICM	rFCM
Lung CT image set	39%	0.46%	0.44%	0.62%	**0.71%**

Conclusion

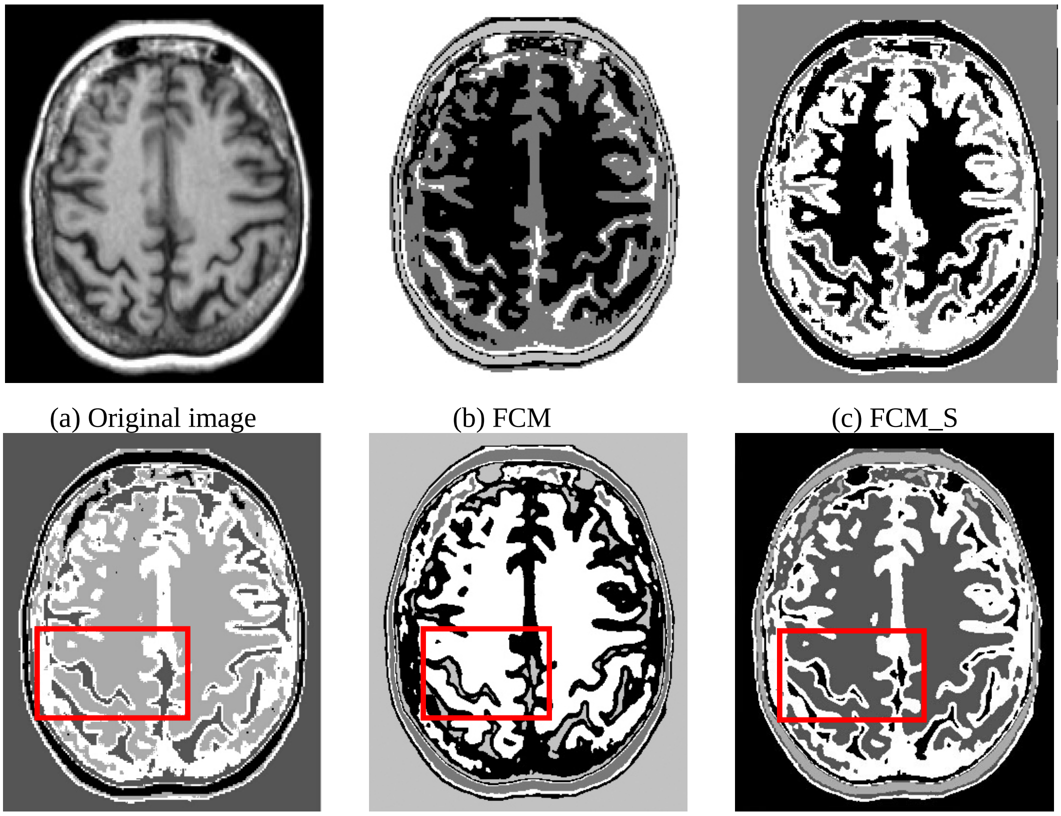
In this paper, we propose a new method based on FCM algorithm, which is applied to the pulmonary cancer Computer-aided Diagnosis (CAD) System. In the segmentation of lung nodules stage, our method (rFCM) depending on characteristics of medical images using neighborhood windows of gray similarity judgments, choose different fuzzy clustering factor for image segmentation process, which solves the traditional algorithm segmentation of images with noise-sensitive issues and FLICM algorithm for image excessive segmentation. Experimental results are applied to the CAD system, which have achieved satisfactory results.



(a) Original image (b) FCM (c) FCM_S

(d) EnFCM (e) FLICM (f) rFCM

Fig. 2 Lung CT image segmentation results (Algorithm parameters $\alpha = 0.7$, $N_R = 9$, $\beta = 6$ and the specification gray level is defined as $gray(N_i)/c$)



(a) Original image (b) FCM (c) FCM_S

(d) EnFCM (e) FLICM (f) rFCM

Fig. 3 Brain CT image segmentation results (Algorithm parameters $\alpha = 0.8$, $N_R = 9$, $\beta = 6$ and the specification gray level is defined as $gray(N_i)/c$)

However, there are still some disadvantages about the new method. For example, it does not work well for the time complexity. In future research, we will try to solve these problems.

Acknowledgement

In this paper, the research was sponsored by the National Natural Science Foundation of China

(NSFC) (Project No. 61572286), the science and technology development plan of Shandong province (Project No.2014GGX101037), Jinan Science and Technology Development Plan (Project No. 201401216).

References

[1] Ahmed, M. N., Yamany, S. M., Mohamed, N., A modified fuzzy C-means Algorithm for bias field estimation and segmentation of MRI data, IEEE Transaction on Medical Imaging, Vol.21, NO.3 (march 2002), pp.193-199.

[2] Ayman, E., Garth, M. B., Georgy, G., Kenji, S., Kazunori, O., Ahmed, E., Ahmed, S., Behnoush, A., Computer-Aided diagnosis systems for lung cancer: challenges and methodologies, International Journal of Biomedical Imaging, Vol.2013(2013) pp.46.

[3] DOI, K., Current status and future potential of computer-aided diagnosis in medical imaging, The British Journal of Tadiology, Vol.78(2005), pp. S3-S19.

[4] Han, F.F., Zhang, G.P., Wang, H.F., Song, B.W., Lu, H.B., Zhao, H., Liang, Z.R., A texture feature analysis for diagnosis of pulmonary nodules using LIDC-IDRI database, Medical Imaging Physics and Engineering(ICMIPE), IEEE International Conference, (2013),pp.14-18.

[5] Jemal, A., Siegel, R., Cancer statistics a Cancer Journal for Clinicians, Vol.58(2008), pp.71-96.

[6] LI, L., Zhang, C.M., Spatio-Temporal Consistency in Depth Video Enhancement, Journal of Advanced Mechanical Design, Systems, and Manufacturing, Vol.7, NO.5(2013), pp.808-817.

[7] Murphy, K., Ginneken, B. V., Schilham, A. M., de Hoop, H.A., Prokop, M., A large scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbor classification, Medical Image Analysis, Vol.13,NO.5(2009),pp.757-770.

[8] Nie, S.D., Sun, X.W., Chen, Z.X., Progress in Computer-Aided Detection for pulmonary nodule using CT image, Chinese Journal of Medical Physics, Vol.26, NO.2(2009), pp.1075-1079.

[9] Samuel, G.A., Geoffrey, M., 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(2011),pp.915-931.

[10] Shen, X.H., Wang, K., Guo, Q., Local thresholding with adaptive window shrinkage in the contourlet domain for image denoising, Sci China Inf Sci, Vol.56(2013), pp.61-69.

[11] Stelios, K., Vassilios, C., A Robust Fuzzy Local Information C-Means Clustering Algorithm, IEEE Transactions on Image Processing, Vol.19, NO.5(2010),pp. 1328-1337.

[12] Szilagyi, L., Benyo, Z., Szilagyii, S.M., Adam, H.S., MR brain image segmentation using an enhanced fuzzy C-means algorithm, Proceedings of the 25th Annual International Conference of the IEEE, Vol.1(2003), pp.17-21.

[13] Wiemker, R., Bergtholdt, M., Dharaiya, E., Kabus, S., Lee, M.C., Agreement of CAD features with expert observer ratings for characterization of pulmonary nodules in CT using the LIDC-IDRI database, Proc.SPIE 7260, Medical Imaging 2009, Computer-Aided Diagnosis, Vol.7260(2009).

[14] Witold, P., James, W., Fuzzy Clustering with Partial Supervision, IEEE Transaction on System, Man, and Cybernetics, Vol.27(1997), pp.1083-4419.