

## **Feature Extraction of Ground-Glass Opacity Nodules using Active Contour Model for Lung Cancer Detection**

Yanli Miao<sup>1, a</sup>, Jianming Wang<sup>1, b</sup>, Weiwei Du<sup>2, c\*</sup>,

Yanhe Ma<sup>3, d</sup>, Hong Zhang<sup>3, e</sup>

<sup>1</sup>School of Electronics and Information Engineering, Tianjin Polytechnic University, Tianjin, 300387, China

<sup>2</sup>Department of Computer Science, Kyoto Institute of Technology, Kyoto, 6068585, Japan

<sup>3</sup>Tianjin Chest Hospital, Nan District, Tianjin, 300051, China

<sup>a</sup>email: [tjpumiaoyanli@163.com](mailto:tjpumiaoyanli@163.com), <sup>b</sup>email: [wangjianming@tjpu.edu.cn](mailto:wangjianming@tjpu.edu.cn),

<sup>c\*</sup>email: [duweiwei@kit.ac.jp](mailto:duweiwei@kit.ac.jp)

<sup>d</sup>email: [mayanhe2008@163.com](mailto:mayanhe2008@163.com), <sup>e</sup>email: [yuxinli72@163.com](mailto:yuxinli72@163.com)

**Keywords:** Feature of Ground-Glass Opacity Nodules; Active Contour Model; Lung Cancer Detection

**Abstract.** The proportion of the solid part portion in a GGO nodule is one of features to detect lung cancer. It is difficult to segment the solid part in a GGO nodule completely because tissues surrounding GGO nodules include some impurities like noises in image processing technology. This paper proposes Active Contour Model (ACM) to find the boundary of a GGO nodule because ACM algorithm can remove noises. The size of a GGO nodule can be computed based on the boundary of the GGO nodule. Expectation-Maximization (EM) algorithm can segment the no solid part in GGO nodules because no solid part and solid part have different densities. Experiments show ACM algorithm is more effective than EM algorithm to find the boundary of a GGO nodule. Moreover, our proposal also can reduce the burden of doctors because it can find the boundary of GGO nodules automatically.

### **Introduction**

In 2016, the American Cancer Society estimates that the mortality rate of lung cancer is the highest in all kinds of cancers. New lung cancer cases will occur in the United States about 224,390 [1]. In Japan, lung cancer has a high mortality rate. There are 77,300 cases [2]. In China, as haze and other environmental pollution reasons, the mortality of lung cancer is up to 610,200 cases in 2015 [3].

A high mortality rate in lung cancer is that survival rate of lung cancer is less than 16% based on the survey in 5 years in all cancer. However, if tumor of lung has been found early, survival rate of lung cancer will be 90% in 5 years [4]. In other words, the more early detection in tumor of lung, the more can improve the survival rate of lung cancer. In lung nodules, there are two types. One is solid nodules, and the other is Ground-Glass Opacity (GGO) nodules [5]. Currently, some approaches [6], [7] have been proposed to detect lung cancer based on some features of solid nodules [8] by image processing methods. However, it is a tough task to detect lung cancer based on GGO nodules because low contrast exists between GGO nodules and tissue surrounding GGO nodules. How to extract features of GGO nodules effectively becomes an important subject in lung cancer detection.

GGO nodules have two types [5]. One is solid part in GGO nodules called part-solid GGO nodules, and the other is no solid part called pure GGO nodules. Part-solid GGO nodules have solid part and no solid part. The sizes of solid part of GGO nodules maybe help doctors to detect malignant tumors. Therefore, we consider that the proportion of the solid part portion in a GGO nodule is one of features to detect lung cancer. The approach of the sizes of solid part of GGO

nodules has been proposed [9]. However, tissues surrounding GGO nodules include some impurities like noises in image processing. [9] Classifies GGO nodules and tissues surrounding GGO nodules based on density by Expectation-Maximization (EM) algorithm. The boundary of a GGO nodule cannot be found very well.

This paper proposes Active Contour Model (ACM) to find the boundary of GGO nodules because ACM algorithm can remove noises. The size of a GGO nodule can be computed based on the boundary of the GGO nodule. EM algorithm can segment the no solid part in the GGO nodule because no solid part and solid part have different density. Experiments show ACM algorithm is more effective than EM algorithm to find the boundary of the GGO nodule. Our proposal also can reduce the burden of doctors because it can find the boundary of GGO nodules automatically.

The remainder of this paper is organized as follows. Section II introduces how to segment GGO nodules with active contour model. Section III explains the algorithm of the feature extraction. Section VI gives some experimental results to certify the effectiveness of our proposal. Section V draws conclusions and future work.

## Feature Extraction with Active Contour Model

ACM can guide external constraint forces and influence image forces to pull an energy spline toward features. The energy spline locks onto nearby edges, localizing them accurately. ACM is applied with detection of edges, lines, and subjective contours, motion tracking and stereo matching [10].

An important characteristic of ACM is to remove impulsive noises for tracking the boundary of objects in computer vision [11]. ACM is considered to have the robustness of filter in image smoothing. The characteristic of ACM is applied to semi-supervised learning [12]. Although the bilateral filter [13] can remove impulse noises, the computational procedure is more complex. This paper applies ACM algorithm to extract feature for detecting lung cancer because ACM can remove impulse noises and find the boundary of a GGO nodule.

ACM is written as the follows:

$$\int_0^1 \mathcal{E}(v(s))ds = \int_0^1 (\mathcal{E}_{\text{int}}(v(s)) + \mathcal{E}_{\text{image}}(v(s)) + \mathcal{E}_{\text{con}}(v(s)))ds \quad (1)$$

Where  $v(s) = (x(s), y(s))$  represents the position in ACM algorithm parametrically,  $\mathcal{E}_{\text{int}}$  represents the internal energy in the spline because of bending,  $\mathcal{E}_{\text{image}}$  represents image forces,  $\mathcal{E}_{\text{con}}$  represents the external constraint forces.

Eq (1) is discretized to  $\int_0^1 \mathcal{E}(v(s))ds = \sum \mathcal{E}_i$  and expressed at eq. (2) depended on [14], [15]. [15] includes hard constraints as described by [16]. Notice that no term for external constraints.

$$\mathcal{E}_i = \alpha \mathcal{E}_{\text{cont},i} + \beta \mathcal{E}_{\text{curv},i} + \gamma \mathcal{E}_{\text{image},i} \quad (2)$$

Where  $\alpha$ ,  $\beta$  and  $\gamma$  are to balance the relative influence of the three terms.

## Algorithm of the Feature Extraction

The ultimate goal of this paper is to find effective features of GGO nodules because the features of GGO nodules can help doctors to find malignant tumor in lung nodules. The algorithm is showed at fig. 1.

The explanation of fig.1 is showed as the followings to the feature of GGO nodules.

(1). to select an image including the biggest size in a GGO nodules from CT volume images. The initial outline is transferred to other images which include the same GGO nodule.

(2). to find the boundaries of all images including the same GGO nodule by using ACM algorithm.

(3). to judge the type of GGO nodules by doctors. The procedure will go to step (5). The size of the pure GGO nodule will be computed if the GGO nodule belongs to pure one.

- (4). to compute the size of solid part in a part-solid GGO nodule by EM algorithm if the GGO nodules are part-solid one.  
 (5). to calculate the size of GGO nodules.

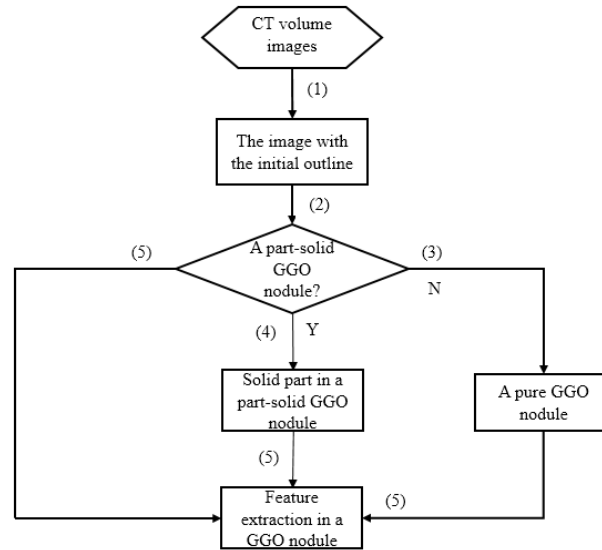


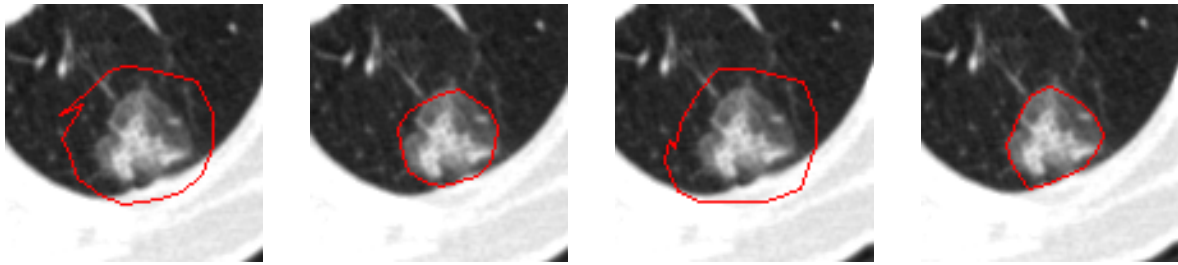
Fig. 1 a flowchart of the feature extraction in a GGO nodule

## Experimental results

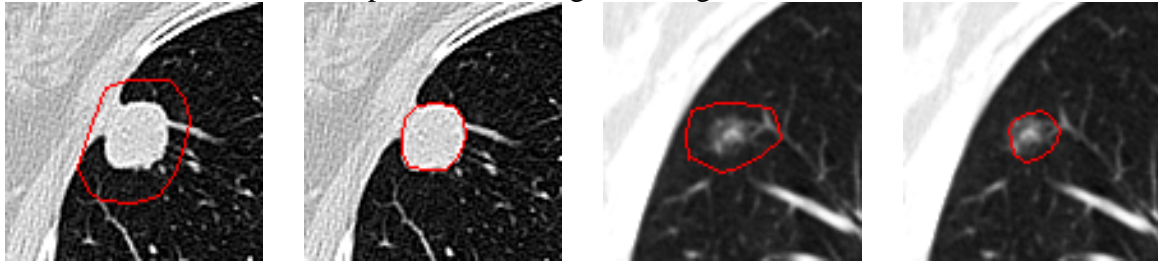
CT volume images are provided with Tianjin chest hospital, China. The scanning range is from the thoracic inlet to the lung bottom. 21 GGO nodules were experimented from 21 patients. The age range of patients is from 26 to 78 years old.

Some parameters of CT volume data are showed as follows: Tube voltage is 120-130KV and tube current is 39-450 mA. Tube voltage and tube current are adjusted depended on the weight of patients. The resolution of one slice is 512 x 512 pixels. The thickness of one slice is 1.5 mm. Only the thickness of one slice to one patient is 5.0 mm because the CT volume data will be given the patient by films form. The range of Field of View (FOV) is 300-446 mm. The FOV is adjusted depended on chest circumferences of patients. The pixel size is 0.59-0.93 mm.

This paper uses the function named `cvSnakeImage` from OpenCV [17] to find the boundaries of GGO nodules. The parameters have been set by  $\alpha=0.15$ ,  $\beta=0.53$ ,  $\gamma=0.016$ ,  $\xi=6$ .  $\alpha$  means the weight of continuous energy.  $\beta$  means the weight of curvature energy.  $\gamma$  means the weight of image energy.  $\xi$  means the number of iterations. Fig. 2 shows the boundaries of the GGO nodules from case 18 with different initial outlines and the same parameters using ACM algorithm. Fig. 3 shows the boundaries of the GGO nodules from case 2 and case 20 with the same parameters. Fig. 3(a) is case 2. The result of case 2 is showed at fig. 3(b). Fig. 3(c) is case 20. The result of case 20 is showed at fig. 3(d). The parameters were set based on experience and can find boundaries of GGO nodules very well.

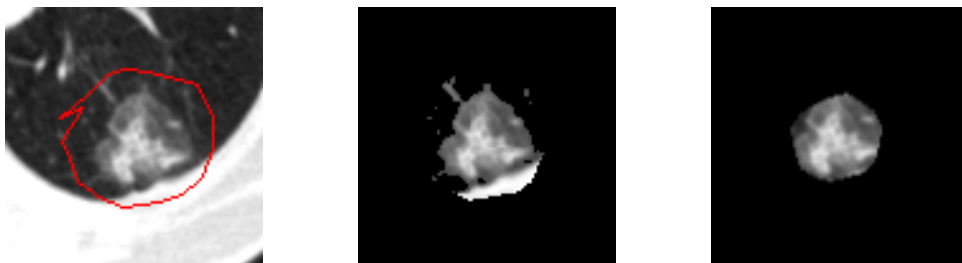


(a). the initial outline (b). the result of (a) (c). another initial outline (d). the result of (c)  
Fig. 2 the boundaries of the GGO nodule from case 18 with different initial outlines and the same parameters using ACM algorithm

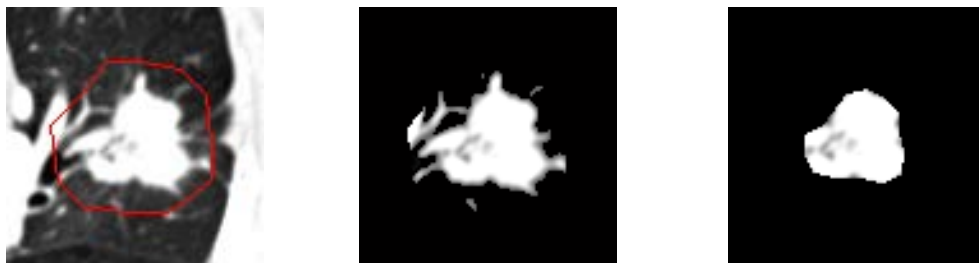


(a). the initial outline (b). the result (a) (c). the initial outline (d). the result (c) by ACM  
Fig. 3 the boundaries of the GGO nodules from case 2 and 20 with the same parameters

The boundaries of the GGO nodule from case 18 were determined using EM and ACM algorithms showed at fig. 4. The GGO nodule cannot be segmented completely using EM algorithm because there are noises in the original image of case 18. Fig. 5(a) is a solid nodule which connects with tissues. The tissues cannot be removed showed at fig. 5(b) using EM algorithm, while the tissues can be removed in fig. 5(c) using ACM algorithm. Table 1 shows the results among ACM, EM and threshold method from case 18. The GGO nodule can be segmented completely using ACM algorithm with visual observation.



(a) the initial outline of case 18. (b) the result of (a) using EM. (c) the result of (a) using ACM.  
Fig. 4 the results of EM and ACM from case 18 with the same initial outline



(a) the initial outline of case 14. (b) the result of (a) using EM. (c) the result of (a) using ACM.  
Fig. 5 the results of EM and ACM from case 14 with the same initial outline

The volume of GGO nodules can be computed based on the boundaries. The volume formula is showed as the followings:

$$V = \sum_{i=1}^N (m \times l^2 \times t) \quad (3)$$

Where  $V$  represents the volume of GGO nodule,  $N$  represents the number of slices which include the same GGO nodule,  $m$  represents the number of pixels of a GGO nodule in a slice,  $l$

represents the spacing between pixels,  $t$  represents the thickness of a slice. Table 2 shows the volumes of GGO nodules and the solid part with 21 cases. The proportion of the solid part portion in a GGO nodule uses the volumes of GGO nodules and the solid part. The formula is showed as the followings:

$$P = \frac{V_{sol}}{V_{GGO}} \quad (4)$$

Where  $P$  represents the proportion of a GGO nodule,  $V_{GGO}$  represents the volume of a GGO nodule,  $V_{sol}$  represents the volume of solid part in a GGO nodule.

Table.2. the volumes and proportions of 21 cases

No. in cases	GGO nodules [mm <sup>3</sup> ]	The solid part [mm <sup>3</sup> ]	The proportion
1	8897.7	5541.2	0.6228
2	2696.1	1997.4	0.7408
3	185.1	60.4	0.3263
4	6161.1	4668.9	0.7578
5	261.0	121.5	0.4655
6	610.6	570.4	0.9342
7	408.3	0	0
8	668.2	385.5	0.5769
9	4903.7	3164.6	0.6453
10	1903.8	1110.0	0.5830
11	5965.8	3658.3	0.6132
12	1497.9	574.9	0.3838
13	1861.4	0	0
14	10321.0	6232.9	0.6039
15	10072.0	7078.6	0.7028
16	52.2	11.9	0.2280
17	3723.9	1836.5	0.4932
18	9952.9	759.8	0.0763
19	1357.7	70.5	0.0519
20	1044.8	0	0
21	10362.0	3848.4	0.3714

## Conclusion

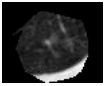
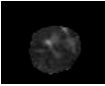

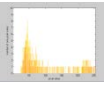

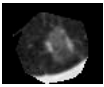
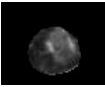
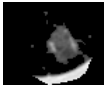
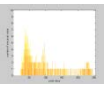




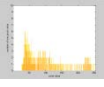




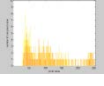




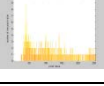






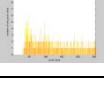







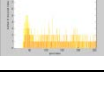























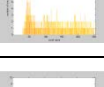


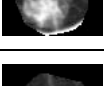




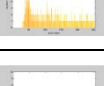


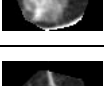







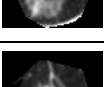



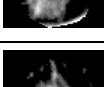



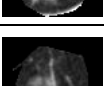
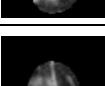
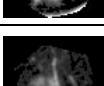

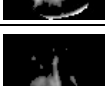
This paper had certified that ACM algorithm can segment GGO nodules completely and remove some tissue surrounding GGO nodules because ACM algorithm has the characteristic of robustness. Experiments showed GGO nodules can be obtained accurately using ACM algorithm. Our proposal can find the boundary of GGO nodules automatically; therefore, it can reduce the burden of doctors. It will be our further work on how to use the proportion feature to help doctor's diagnosis.

## References

- [1] Rebecca L. Siegel, Kimberly D. Miller, and Ahmedin Jemal, Cancer Statistics, 2016[J], CA: A Cancer Journal for Clinicians, 2016 66 (1) 7-30.
- [2] Katanoda K, Kamo K I, Saika K, et al. Short-Term Projection of Cancer Incidence in Japan

- Using an Age-Period Interaction Model with Spline Smoothing[J]. Japanese Journal of Clinical Oncology, 2014 44(1) 36-41.
- [3] Wanqing Chen, Rongshou Zheng, Peter D. Baade, et al, Cancer statistics in China, 2015[J], CA: A Cancer Journal for Clinicians, 2016 66 (2) 115-132.
- [4] C I Henschke, DI Mccauley, DF Yankelevitz, et al. Early Lung Cancer Action Project: Overall Design and Findings from Baseline Screening[J]. Lancet, 2000 354(9173) 99-105.
- [5] Hee-Dong Chae, Chang Min Park, Sang Joon Park, Sang Min Lee, Kwang Gi Kim, Jin Mo Goo. Computerized Texture Analysis of Persistent Part-Solid Ground-Glass Nodules : Differentiation of Preinvasive Lesions from Invasive Pulmonary Adenocarcinomas[J], Radiology, 2014 273(1) 285-293.
- [6] Omnia Elsayed, Khaled, Mahar, Mohamed Kholief, et al. Automatic Detection of the Pulmonary Nodules from CT Images[C]. SAI Intelligent Systems Conference. UK: London, 2015 742-746.
- [7] Atsushi Yaguchi, Tomoya Okazaki, Tomoyuki Takeguchi, et al. Semi-Automated Segmentation of Solid and GGO Nodules in Lung CT Images using Vessel-Likelihood Derived from Local Foreground Structure [C]. Proc. of SPIE. Japan: 2015 742-746.
- [8] Heber MacMahon, John H. M. Austin, Gordon Gamsu, et al. Guidelines for Management of Small Pulmonary Nodules Detected on CT Scans: A Statement from the Fleischner Society[J], Radiology, 2005 237(2) 395-400.
- [9] Yanli Miao, Jianming Wang, Weiwei Du, Yanhe Ma and Hong Zhang. Volumetric Measurement of Ground-Glass Opacity Nodules using Expectation-Maximization Algorithm[C]. The 4th IIAE International Conference on Intelligent Systems and Image Processing 2016: The Institute of Industrial Applications Engineers. Japan: Kyoto, 2016 317-321
- [10] Kass M., Witkin A., and Terzopoulos D., Snakes: Active Contour Models[J]. International Journal of Computer Vision, 1988 321-331.
- [11] Chan B T F, Vese L A. Active Contours without Edges, 2001[C]. 2007 IEEE International Conference on Robotics and Automation, 2007. 2012.
- [12] Weiwei Du, Yi-Peng Liu. Semi-Supervised Learning Spectral Embedding with Active Contours Model[C]. IEEE Conference Publications. 2015 797-801.
- [13] Garnett R, Huegerich T, Chui, C, Wenjie He. A Universal Noise Removal Algorithm with an Impulse Detector[J]. IEEE Transactions on Image Processing. 2005 14(11) 1747-1754.
- [14] Amini A A, Tehrani S, Weymouth T E. Using Dynamic Programming for Minimizing the Energy of Active Contours in the Presence of Hard Constraints[C]. International Conference on Computer Vision, ICCV 1988. Tampa, Florida, Proceedings. 1988 461-478.
- [15] Williams D J, Shah M. A Fast Algorithm for Active Contours and Curvature Estimation[J]. CVGIP: Image Understanding, 1992 55(1) 14-26.
- [16] X. Xie, R. Sudhakar and H. Zhuang. On Improving Eye Feature Extraction using Deformable templates[J]. Pattern Recognition: The Journal of the Pattern Recognition Society. 1994 27(6) 791-799.
- [17] Gary Bradski, Adrian Kaehler, Learning OpenCV: Computer Vision with the OpenCV Library [M]. USA: O' Reilly Media, Inc, 2008

Table 1. Comparison among ACM, EM and threshold method

No. in slices	Images with initial outline	ACM		EM		Histogram	A threshold method	
		Solid part	GGO nodules	Solid part	GGO nodules		Solid part	GGO nodules
82								
83								
84								
85								
86								
87								
88								
89								
90								
91								
92								
93								
94								
95								
96	