

## 3D PULMONARY NODULES DETECTION USING FAST MARCHING SEGMENTATION

M. P. Paing and S. Choomchuay\*

Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok,  
Thailand

Published online: 05 October 2017

### ABSTRACT

Pulmonary nodule detection is an important step in lung cancer detection because nodules are the alert signal of lung cancer. The early detection of them can hence increase the patient's survival rates. This paper proposes an automated computer aided diagnosis system for detection of pulmonary nodules based on three dimensional (3D) structures. Lung parenchyma segmentation using fast marching method was employed. A simple thresholding technique is used to extract candidate nodules from the segmented lung parenchyma. A 3D image of nodule candidates is reconstructed by mean of stacked 2D images. To find the connected voxels of a blob, a 3D connected component labelling is used. Features extracted from each blob are then fed into the classifier. The random forest algorithm has been invoked for nodule and non-nodule classification. The proposed detection methodology can give the accuracy of 92%.

**Keywords:** lung cancer; pulmonary nodule; fast marching; 3D features; random forest classifier.

Author Correspondence, e-mail: [somsak.ch@kmitl.ac.th](mailto:somsak.ch@kmitl.ac.th)

doi: <http://dx.doi.org/10.4314/jfas.v9i4s.19>



## 1. INTRODUCTION

The fight against cancer has been conducted as a global issue intended to save the valuable human lives. The World Health Organization (WHO) accelerates awareness and education about cancer in four basic components of cancer control-prevention, early detection, diagnosis and treatment, and palliative care. Many cancers can be prevented by detecting at the early stage of cancer development. As such, early detection helps cancer patients to get higher potential to cure and effective treatment can increase the survival rate of cancer. The patients' awareness of the early symptoms of the disease and consulting with the physician is a kind of early detection [1]. However, most of the cancer rarely shows the symptoms at the early stage of the disease. This fact is the main reason leading the cancer death.

Computer aided diagnosis systems (CAD) have played an important role in early detection of the diseases for decades. The CADs are computerized systems aiming to reveal specific abnormalities during the interpretation of patient radiology images [2]. The results of CAD are used as a second option for diagnostic decision. The CADs systems are widely used in lung cancer detection for early diagnosis because most of the lung cancers do not show any noticeable symptoms until they have spread, but some patients with early lung cancer do have symptoms [3]. Pulmonary nodules, the spots having 1mm- 30mm in diameters on the Computed Tomography (CT) are the most common alert signal of possible lung cancer. Detection the cancerous pulmonary nodules as early resection of stage can increase the 5-year survival rates as high as 70% [4].

Many literatures have been published on developing CADs for lung cancer detection. Review of literatures indicates that pulmonary nodule extraction can be conducted in three main processes namely the lung parenchyma extraction, lung border reconstruction and nodule candidates detection.

The first process, lung parenchyma extraction, is performed at the earliest stage. This is because the region of interest, pulmonary nodules, are located in this area. As a result, different segmentation methods were proposed. Among those, segmentation using thresholding and morphology was the most common technique for lung parenchyma segmentation [5-6]. Although such a technique is fast and easy, the selection of appropriate

---

thresholding value is the most important one. The iterative optimal thresholding was proposed subsequently to solve the threshold value problem [7]. The prior probabilities, and object and background distribution limitation must be known. In other researches [8-9], marker control watershed segmentation was also used to segment the lung parenchyma. However, in some cases this method is sensitive to noise and can produce the over segmentation result. Active contour [10-12] was an alternative method of lung parenchyma extraction process. It can successfully separate the foreground objects from background in both two dimensional (2D) and three dimensional (3D) images. Nevertheless, segmentation using active contour takes long computation time because of the iterative process. Moreover, it has the weakness that is quite sensitive to noise and less sensitive to weak edge [13]. The second process in pulmonary nodule extraction, lung border reconstruction, is performed especially for some nodules which are attached to the lung wall. These nodules are so known as Juxta-pleural. Morphological operation, rolling ball algorithm and chain code analysis are generally used for border reconstruction [14].

The third process, detection of nodule candidates, is performed generally by superimposing lung parenchyma mask onto the original image and thresholding [6-7, 9].

Be noted that the techniques described above were all done in traditional 2D segmentation. As CT scanning produces multiple slices from different dimensional views, extracting the nodules' features only from the 2D view cannot provide sufficient classification results. For this reason, 3D image processing methods for nodule segmentation were introduced in later researches. In [15], a pulmonary nodule detection scheme utilizing 3D feature extraction and support vector machine (SVM) classification was proposed. They segmented lung parenchyma using bounding box and binarization. Then, they applied rolling ball technology and dot enhancement filter for segmentation of initial nodule candidates. Another automatic method for nodule segmentation using 3D visualization technique, active contour modeling and KNN classifier was described in [16]. Moreover, 3D multi-scale Block LBP filter for nodule candidates segmentation was proposed in [17].

With the improved performance of 3D, the computational effort also increases because 3D image processing is a bit more complex than 2D and it also needs greater processing time.

Hence, in [18] proposed a combination method of 2D segmentation and 3D feature extraction. In their study, a 2D segmentation algorithm using thresholding was applied to segment lesions inside the lung parenchyma. The proper threshold value was chosen based on both histogram and nodule size. The nodule size of 4 mm was taken as a limit since the smaller size was considered as 0% malignancy. 3D lung reconstruction was then performed and followed by 3D blob detector algorithm for candidate nodule extraction. Then 3D features were extracted and nodules were classified. Support vector machine (SVM) was employed as a classifier.

The major contribution of our research work is to develop a system that can detect the pulmonary nodules using 2D and 3D hybrid techniques. This is also similar to the work proposed by [18] that segmentation is carried out in 2D environment whilst feature extraction is performed in 3D environment. In difference and extras, this study has proposed three main contributions; (i) a new segmentation method for lung parenchyma extraction, (ii) 3D connected component labelling and (iii) random forest classifier. The fast marching method for lung region segmentation is introduced. In this method, seed locations are defined by a global threshold value while pixel weights are calculated based on the difference between each pixel's intensity and the average intensity of all pixels. Although the fast marching technique introduces these additional steps, it has eased the selecting of a proper threshold value that generally required in [18]. In additions, edge detection and morphological operations were borrowed to solve the border reconstruction problem which is always a case in juxta-pleural nodule segmentation. Furthermore, 3D component labelling is simpler and easier than 3D blob detector used because the blob detector can be sometimes confused especially when the vessels branch as a blob.

There are several classifiers used to automate object classification, e.g. fuzzy, neural network, K-nearest neighbor, principal component analysis, support vector machine and so on. In this work, we considered the random forest classifier because it can improve the classification accuracy by using ensemble multiple decision trees. Although SVM in [18] can give the high accuracy of classification, it depends on the optimal choice of kernel.

## 2. METHODOLOGY

The overview of the proposed system shown in Fig. 1 consists of five main stages: input image (materials), segmentation, 3D nodule candidates reconstruction, feature extraction and classification.

- Input image: The lung CT scan images are to be inputted into the system.
- Segmentation: Refers to the process of separating the region of interest (ROI) from the other parts of the image. In this stage, the lung parenchyma region which includes gas, vessels, fissures and nodules is segmented first. Then, the objects called nodule candidates which have a higher probability to be nodules are extracted from each two dimensional CT slice.
- 3D nodule candidate reconstruction: Creates a 3D image of nodule candidates by stacking the extracted nodule candidates from all 2D slices of any single case.
- Feature extraction: 3D features are extracted from nodule candidates.
- Classification: Classification of nodule and non-nodule is performed using random forest classifier.

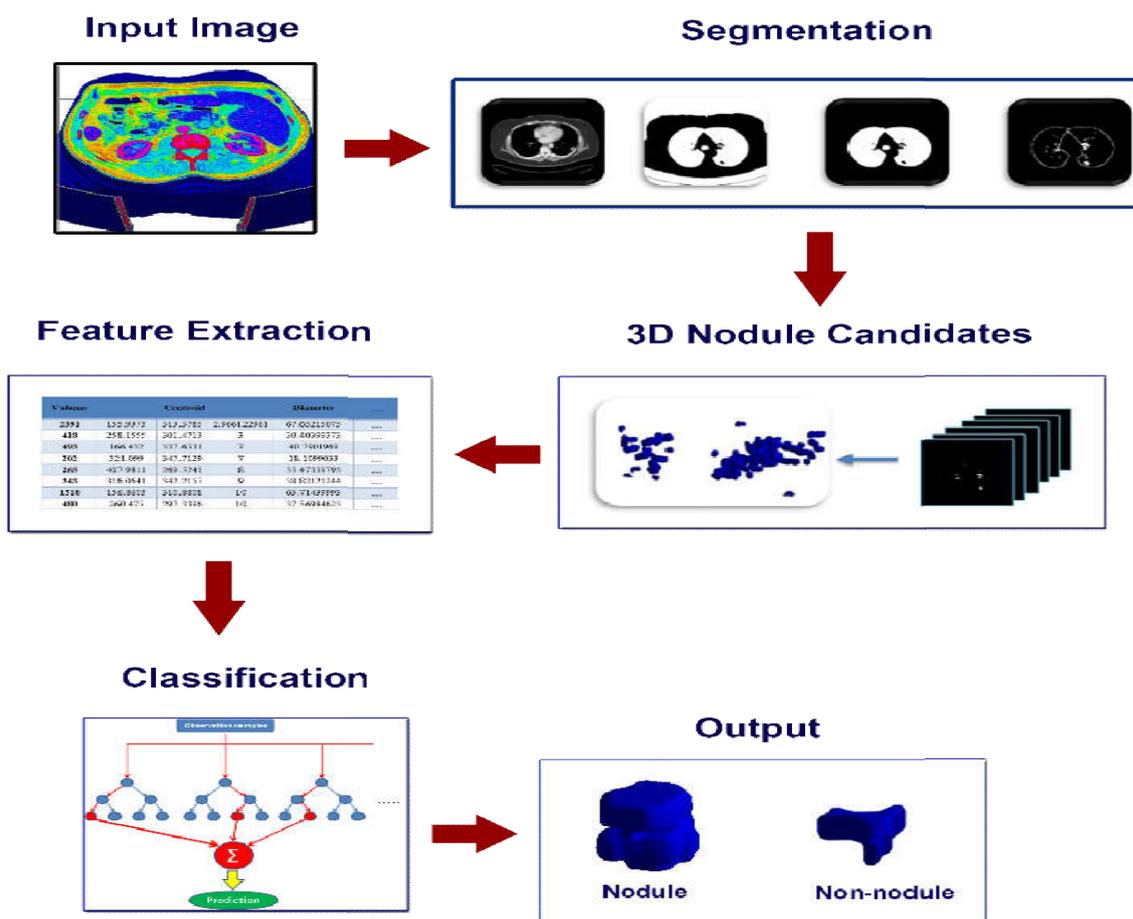


Fig.1. Outline of proposed methodology

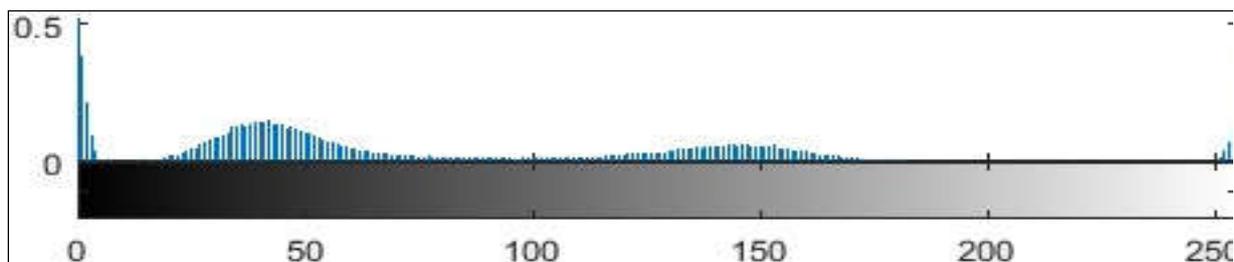
## 2.1. Materials

LungCT-Diagnosis collection from TCIA [19] is used as an input dataset in this research. The collection is originally provided by the Moffitt Cancer Centre (Tampa Florida) [20-21]. The downloaded CT slices are in DICOM format that slice thickness are varied from 3 to 6 mm. The collection also provides the representative tumour slices file describing the instance number of the tumour slices. Before inputting into the system, slices are converted into the JPEG format. The representative tumour slices file is used as the ground truth in the experiment.

## 2.2. Segmentation

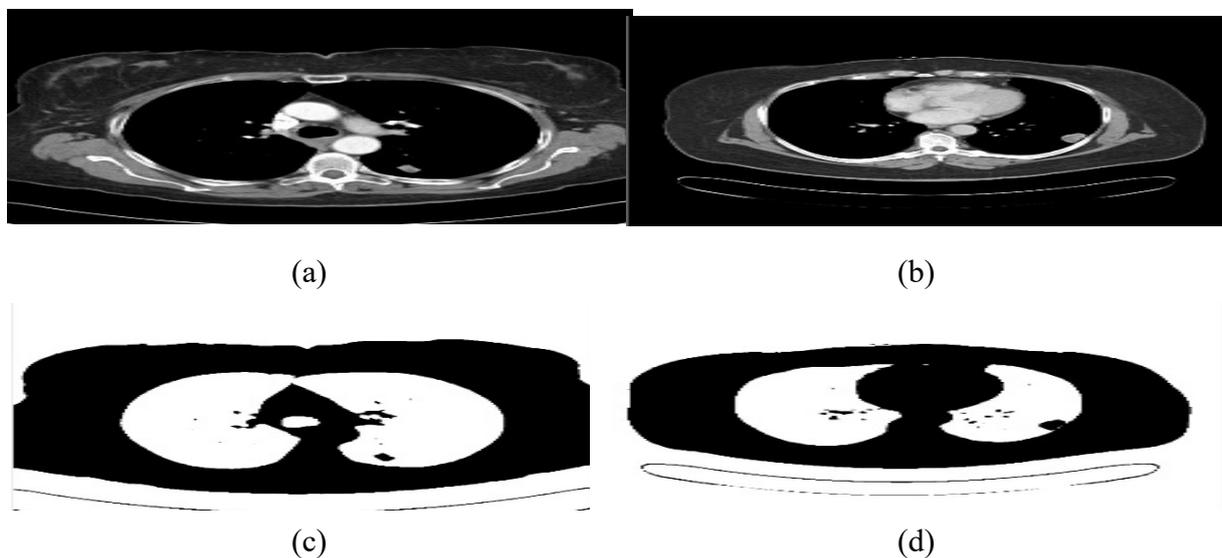
This stage tries to segment the nodules from the converted input images. A method based on the fast marching algorithm has been considered. The algorithm is originally aimed to solve the boundary value problem with two given initial conditions [22-23]. It is a computationally efficient method for tracking an evolving contour [13]. Fast marching based segmentation starts by roughly defining the region of interest (ROI) with two contours, inner and outer contours. The pixels belonging to each contour are assigned with different labels. To find the edges of the targeted object, the algorithm propagates these two initialized contours into opposite directions; the inner to outside and the outer to inside. The boundary of ROI is where the two contours contact. The propagation stops automatically when the two contours converge over each other.

To perform the segmentation, the seed locations are defined by a mask created based on the global image threshold value using Otsu's method [24]. Such a method holds the desirable feature that it can separate foreground and background objects effectively. This is particularly true for an image with a bimodal histogram otherwise it makes additive noise on the image. The histogram of one sampled CT slices is shown in Fig. 2. Although the shown histogram is not a perfect bimodal, Otsu's method is still applicable because thresholding method is not employed for segmentation. In contrast, the obtained threshold value is required for finding seed location. In our case, the best threshold value obtained from Otsu's method is 0.3.



**Fig.2.** Histogram of an input CT slice

A false logical array as the same size with the input image is firstly defined as a mask. All pixels in the mask are then compared with the global threshold value. Those with smaller value are converted to true. The true locations in the mask are the seed locations. The pixel weights are computed from the difference between the intensity of the pixel and the average of the intensity values of all the pixels in the mask. The high weight values are segmented as the foreground object and the low values as the background. Fig. 3 shows the examples of input CT slices; (a) a nodule inside the lung parenchyma, (b) a Juxta-pleural nodule (a nodal attaches to the lung wall). (c) and (d) are the result of (a) and (b) after fast marching segmentation.



**Fig.3.** Examples of input CT slices (a) Nodule inside the lung parenchyma, (b) Juxta-pleural nodule, (c) Segmented Image of (a), (d) Segmented Image of (b)

After separating the foreground and background objects, border clearing is conducted in order to get the lung parenchyma region where nodules reside. Fig. 4 shows the lung parenchyma region.

Then, edge reconstruction of the lung parenchyma is performed because the edges of the lung parenchyma region with juxta-pleural nodules are disconnected. A conventional edge detection method is applicable. As shown in Fig. 5 (b), the edge of the lung parenchyma is disconnected at the location where the Juxta-pleural nodule exists.



**Fig.4.** Lung Parenchyma Region (a) Nodule inside the lung parenchyma and (b) Juxta-pleural nodule

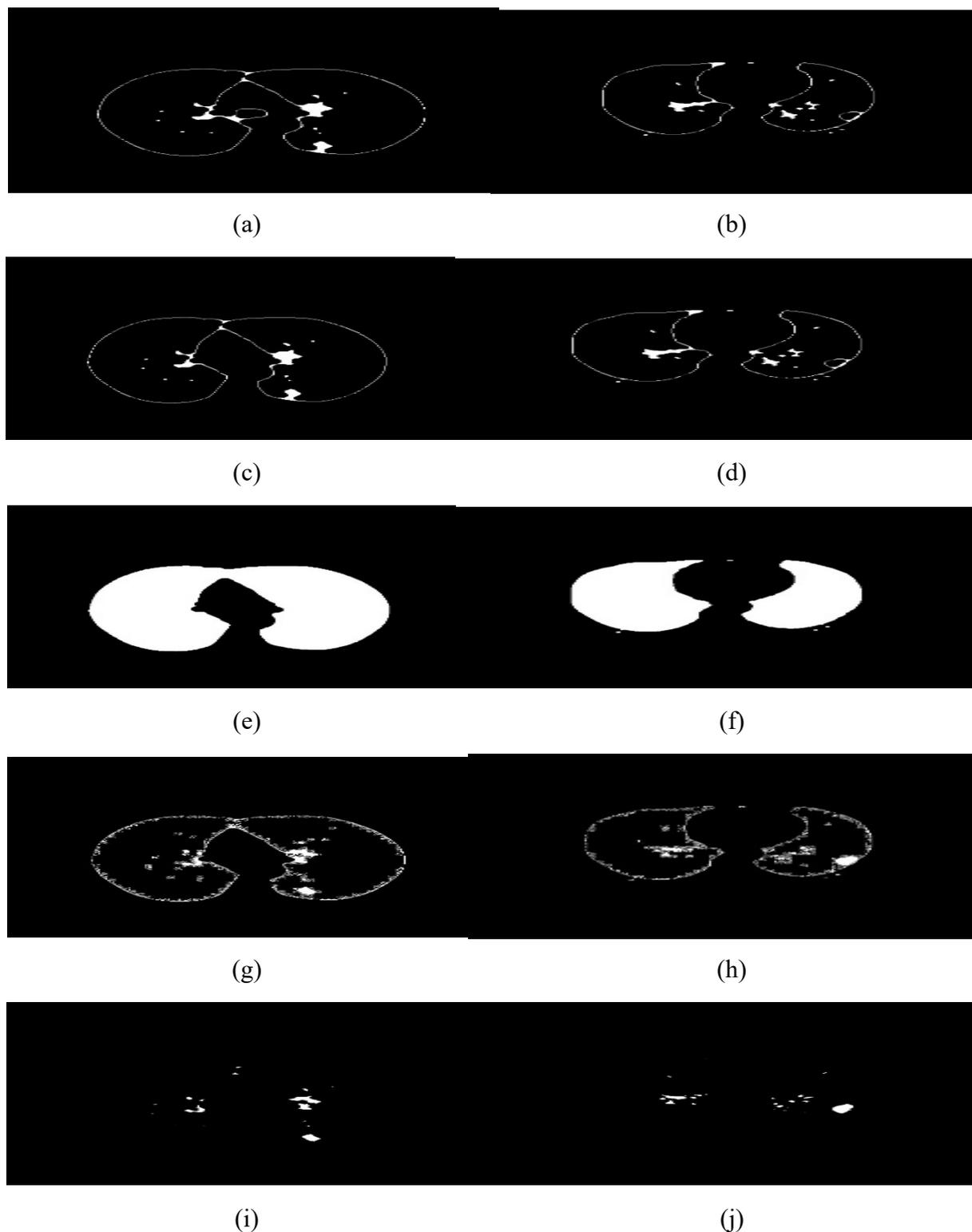


**Fig.5.** Edge of the Lung Parenchyma region (a) Nodule inside the lung parenchyma and (b) Juxta-pleural nodule

Morphological closing with a disc shaped structuring element of size 6 is used to connect the edge as shown in Fig. 6 (a) and (b). However, the closing causes the undesirable areas by connecting the separate edges. These undesirable objects can increase false positives in detection. Therefore, some extra steps are required to reduce the false.

As illustrated in Fig. 6 (a), the undesirable pixels appear by connecting the edges of lung parenchyma and the trachea, the circle- shaped in the middle of the image. Since only the lung parenchyma is the place where the nodules (ROI) locates, the deleting the trachea cannot affect the ROI detection. Therefore, the trachea is deleted before using the closing effect. Fig. 6 (c) and (d) show the image after deleting the trachea.

After deleting the trachea, the hole filling method is applied to get the lung parenchyma masks as shown in Fig. 6 (e) and (f). The resulted lung parenchyma masks are superimposed onto original images Fig. 3 (a) and (b). This can give successfully segmented image of the lung parenchyma regions as shown in Fig. 6 (g) and (f). Then, thresholding of the segmented parenchyma images finally produces the segmentation of the lesions inside the lung parenchyma.



**Fig.6.** Segmentation of the nodule candidates (a, b) Closing the parenchyma edge (c, d) Deleting the trachea (e, f) Filling the holes in the parenchyma region (g, h) Segmented parenchyma (i, j) Segmented lesions inside the lung parenchyma

Obtaining all the lesions inside the lung parenchyma region, a selection of the nodule candidates is then conducted because the lung parenchyma includes not only the nodules but

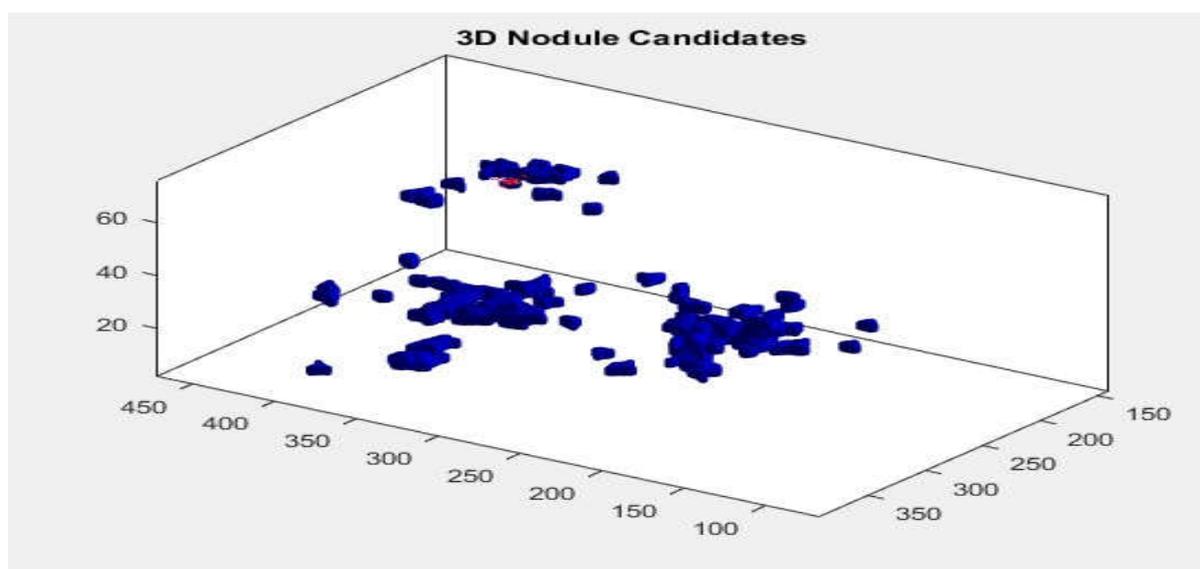
also other lesions such as vessels, gases and fissures. Nodules are rounded opacity on the CT scan having the size arranging from 1mm to 30 mm. Therefore, the lesions that are smaller than 1mm and line shaped objects are eliminated in order to leave only the objects that have the possibility to be nodules. These objects are called nodule candidates and depicted in Fig. 7.



**Fig.7.** Nodule candidates (a) Nodule inside the lung parenchyma and (b) Juxta-pleural nodule

### 2.3. 3D Nodule Candidate Reconstruction

Detection the nodules from a single 2D slice is not enough to produce the accurate diagnosis result because the CT scanning produces hundreds of 2D slices of the patient's lung from multiple dimensions. Nodules may disappear or may be small in some slices. In the real world, the radiologists manually read these hundreds of slices to make a final diagnosis result. It is a time and effort consuming task and sometimes it may lead to human errors. This regard brings in the computer aided system. The nodule candidates in each 2D CT slice of a single case are stacked and reconstructed to get a 3D structure as shown in Fig. 8.



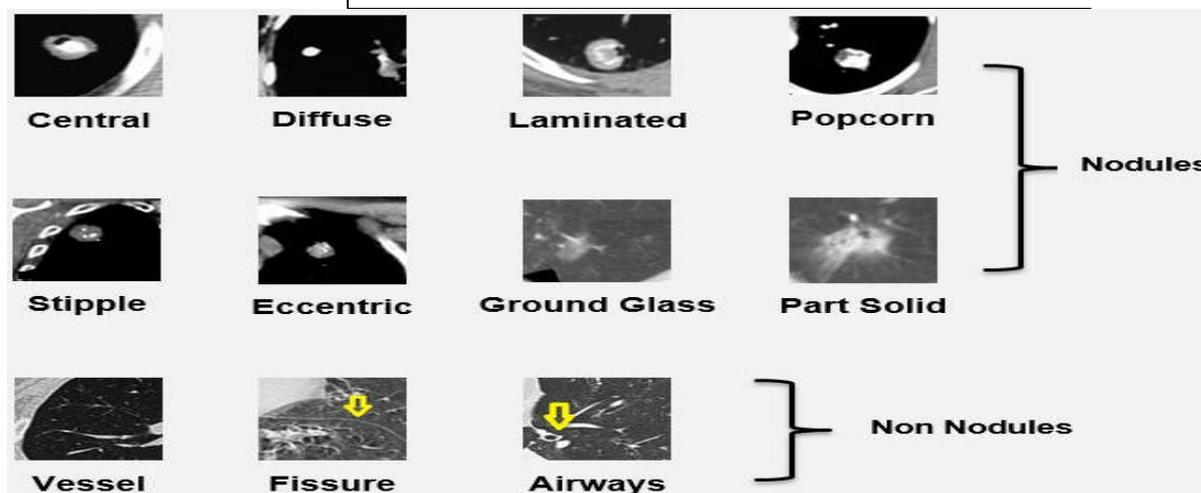
**Fig.8.** Example of the 3D image of nodule candidates of a single case

### 2.4. Feature Extraction

Before extracting the 3D features from the nodule candidates, the 3D connected component labelling is performed to know which voxels are clustered as a blob. Connected component labelling algorithm finds all connected components in an image and assigns a unique label to all points in the same component. The connectivity 26 is used for three dimensional images. Connected component labelling outputs a label matrix with the same size as the input image. The elements in the label matrix are integer values greater than or equal to 0. The background is labelled as 0 and blobs are labelled in numerical orders such that the first blob is labelled as 1, the second blob is labelled as 2 and so on.

**Table 1.** Radiological criteria of nodules [25]

<b>Shape</b>	Focal, rounded, polygonal opacity
<b>Size</b>	≤ 3 cm in diameter
<b>Calcification</b>	Central, diffuse, popcorn, laminated, stipple, eccentric
<b>Margin</b>	Smooth, Irregular
<b>Density</b>	Homogeneously solid or ground glass Heterogeneously solid or ground glass



**Fig.9.** Differences between nodules and non-nodules [26-30]

3D features of each blob are then extracted and fed into the classifier. Features are extracted to meet with the radiological features of pulmonary nodules. Therefore, the radiological criteria which are generally used to detect the nodules in the real world are studied first. According to [25], radiological criteria of the pulmonary nodules are described in Table 1. Moreover, the main differences between the nodules and non-nodules structures are depicted in Fig. 9

[26-30].

In this proposed research, the 3D geometrical features such as the eccentricity, volume, diameter, surface area, margin and density are extracted to classify the shape, size and density of the pulmonary nodules.

**Eccentricity:** It is a scalar value which returns the ratio of the distance between the foci and major axis length of the blob. The value of eccentricity is between 0 and 1. The value 0 indicates a circle and 1 indicates a line segment. It is calculated by:

$$\text{Ecc} = \frac{f}{\text{majoraxislength}} \quad (1)$$

where Ecc means the eccentricity of the labelled blob and f means the foci of the blob.

**Volume:** It is a scalar value which defines the space that the blob takes up. It counts the actual number of voxels in the labelled blob. It is calculated by:

$$V = \sum (x_l, y_l, z_l) \quad (2)$$

where V is the volume of the labelled blob and x, y and z are the coordinates of the voxels of the labelled blob l.

**Diameter:** It is a scalar value that is the greatest possible distance of a line through the center of the blob. It is calculated by:

$$d = 2 * \sqrt[3]{\frac{3V}{4\pi}} \quad (3)$$

where d is the diameter of the blob and V is the volume of the blob.

**Surface Area:** It is a scalar value that defines the total number of pixels on the outside surface of the labelled blob. It is calculated by:

$$\text{surface area} = 4\pi r^2 \quad (4)$$

where r is the radius of the labelled blob and it is the half of the diameter.

**Margin:** It is a scalar value that defines the soft or irregular margin of the blob. To calculate the margin, the circumference of the blob is computed first. It is calculated by multiplying the diameter with  $\pi$  (pi). Then, the ratio of the circumference to the surface area is calculated to get the margin. The larger margin value shows an irregular margin of the blob.

$$\text{margin} = \frac{C}{A} \quad (5)$$

where C is the circumference and A is the surface area of the blob.

**Density:** It is a scalar value that shows the density of the voxels in the labelled blob. It is a ratio of the volume which is the number of 'on' voxels to the total number of voxels in the blob.

$$\text{density} = \frac{V}{\text{numel}(l)} \quad (6)$$

where V is the volume, numel(l) is the number of array elements in the labelled blob.

Moreover, the texture features such as the correlation, contrast, entropy, energy and homogeneity are also extracted using 3D gray level co-occurrence matrix (GLCM). Fig. 10 shows the structure of the 3D GLCM [31] which is an extension of 2D GLCM, the most useful method for textured analysis introduced by [32].

**Correlation:** It is the measurement of how a pixel is correlated to the neighborhood pixels. Correlation outputs the value between -1 and 1. The value -1 means perfectly negatively correlated image and 1 means positively correlated image.

$$\text{correlation} = \sum_{i,j,k} P_{ijk} \frac{(i - \mu_i)(j - \mu_j)(k - \mu_k)}{\sigma_i \sigma_j \sigma_k} \quad (7)$$

**Contrast:** It is the measurement of the local variations in the gray-level co-occurrence matrix. It returns the difference between the highest and the lowest values of a contiguous set of pixels.

$$\text{contrast} = \sum_{i,j,k} P_{ijk} [(i - j)^2 + (i - k)^2 + (j - k)^2] \quad (8)$$

**Entropy:** It is a measurement of the disorder or complexity of an image. The large entropy value means that the image is not uniform in texture and the GLCM elements in this image are very small, i.e. high entropy means complex textures.

$$\text{entropy} = \sum_{i,j,k} P_{ijk} (-\ln P_{ijk}) \quad (9)$$

**Energy:** It is a measurement of textural uniformity and also known as uniformity or the angular second moment. It detects disorders in textures by summarizing the squared elements in the GLCM. It ranges between 0 and 1.

$$\text{energy} = \sum_{i,j,k} P_{ijk}^2 \quad (10)$$

**Homogeneity:** It is a measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. It also ranges between 0 and 1.

$$\text{homogeneity} = \sum_{i,j,k} \frac{P_{ijk}}{1 + [(i - j)^2 + (i - k)^2 + (j - k)^2]} \quad (11)$$

In the above formule, p(i, j, k) is the (i, j, k)<sup>th</sup> pixel values in GLCM and  $\mu_i$ ,  $\mu_j$ ,  $\mu_k$ ,  $\sigma_i$ ,  $\sigma_j$  and  $\sigma_k$  are the means and standard deviations of  $p_i$ ,  $p_j$  and  $p_k$ .

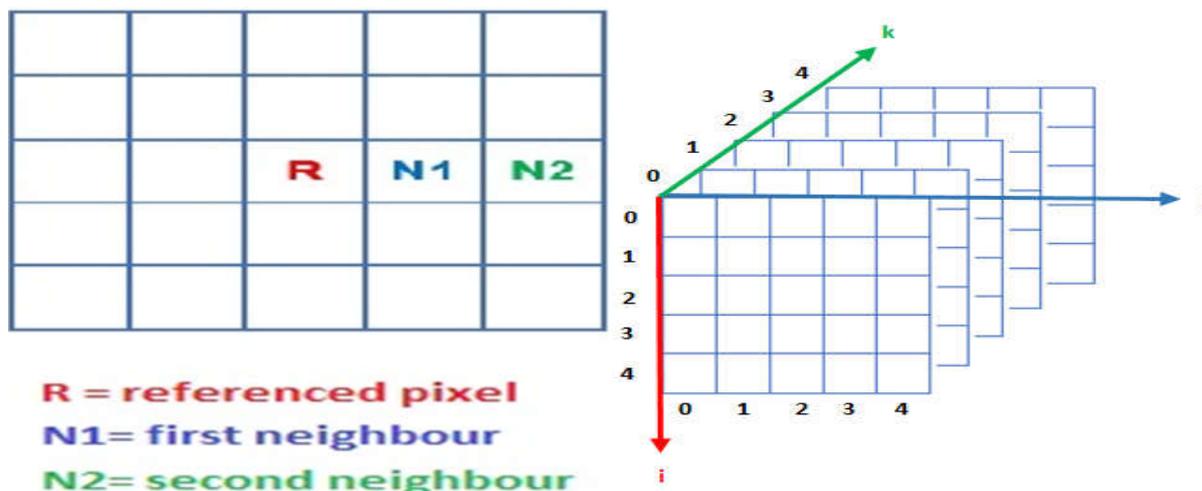


Fig.10. Structure of 3D GLCM [31]

### 2.5. Classification

Random forest classifier illustrated in Fig. 11 [33] is an ensemble learning method that performs the classification by training lot of decision trees on random features with corresponding classes. In random forest, a decision tree classifier is applied on each random subset of data and produces the predicted classes. The decision trees which are used in the random forest do not interact with each other and execute in parallel. The final output class of the random forest is chosen by finding the mode output class of the individual tree. Ensemble algorithm called Bootstrap aggregating or bagging is used for training algorithm of the random forest. The main difference between the original bagging tree and random forest is that random forest uses a random subset of features. The main reason for this is that most decision trees can provide the correct prediction of class for the most part of the data and makes mistakes in different places.

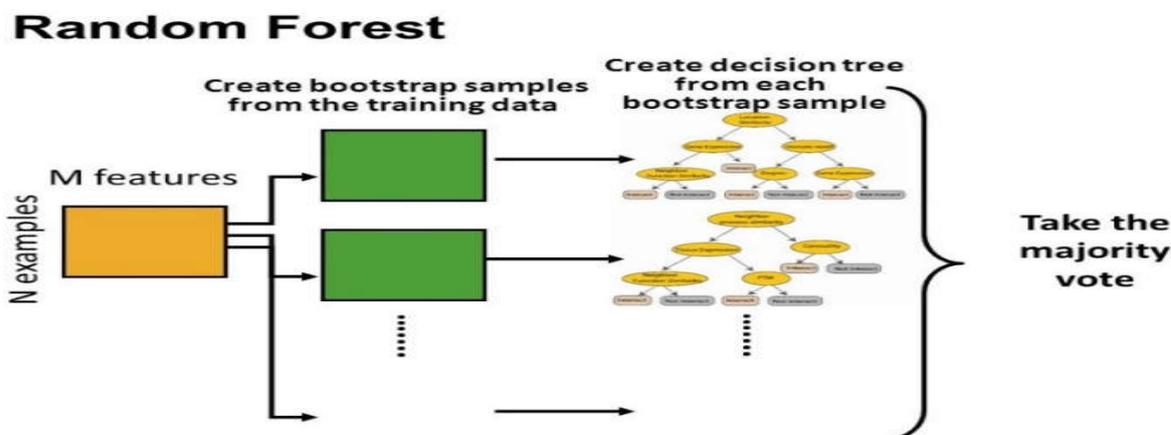


Fig.11. Structure of the Random Forest Classifier [33]

Although the random forest can give the correct prediction, overfitting which may lead to

wrong prediction can be occurred when the trees are over grown. It can be prevented by optimizing a tuning parameter which governs the number of random features to grow each tree. In this proposed research, a random forest classifier with 100 decision trees is used to classify the nodules and non-nodules. Each tree has 9 depths and 5 folds are used to train each weak learner.

### 3. RESULTS AND DISCUSSION

The experiment is conducted using 61 different cases with 3870 total slices downloaded from LungCT-Diagnosis collection. Each case consists of about 60 to 150 slices and single nodule per case. 30 cases of the collection are used for training and 31 cases are used for testing. The random forest classifier with 100 decision trees is used for nodules and non-nodules classification. The classifier is trained using eleven extracted features from segmented nodules. The major voting method is applied to select the final prediction from 100 trees.

Table 2 describes the example of the dimension of the extracted feature vectors and the final classification result from the random forest classifier. In this example, there are 6 candidate nodule blobs referred to as B-1 to 6.

The prediction results obtained from the confusion matrix are used to evaluate the performance of the system. The four statistical measures of the performance such as sensitivity, fall-out, precision and accuracy are calculated for performance of the system by using the formulae described below.

$$\text{Sensitivity} = \frac{\sum \text{True Positive}}{\sum \text{Condition Positive}} \quad (11)$$

$$\text{Fall - out} = \frac{\sum \text{False Positive}}{\sum \text{Condition Negative}} \quad (12)$$

$$\text{Precision} = \frac{\sum \text{True Positive}}{\sum \text{Test Outcome Positive}} \quad (13)$$

$$\text{Accuracy} = \frac{\sum \text{True Positive} + \sum \text{True Negative}}{\sum \text{Total Population}} \quad (14)$$

**Table 2.** An example of the features extracted and classification results

Features	B-1	B-2	B-3	B-4	B-5	B-6
Eccentricity	1	0.9	1	1	1	1
Volume	289	236	246	351	1442	1338
Diameter	25.4	17.5	29.5	29.7	53.1	63.3
Surface Area	574.9	382.8	495.8	693.1	2707	2557
Margin	0.14	0.16	0.18	0.13	0.06	0.07
Density	1.6	1.3	1.3	1.9	7.9	7.5
Correlation	-0.03	-0.45	-0.19	-0.17	-0.06	-0.46
Contrast	3638	6941	6487	3376	10770	7606
Entropy	5.66	9.53	5.50	5.86	7.25	7.19
Energy	0.003	0.01	0.004	0.002	0.0007	0.0007
Homogeneity	0.0003	0.0005	0.0001	0.0003	0.0001	0.0001
Final Result	Non-nodule	Non-nodule	Nodule	Non-nodule	Non-nodule	Non-nodule

The proposed system achieves 96% sensitivity, 12.5% fall out, 88% precision and 92% accuracy. The confusion matrix of the testing is given in Fig. 12.

	Nodule	Non-nodule
Nodule	30	1
Non-nodule	4	28

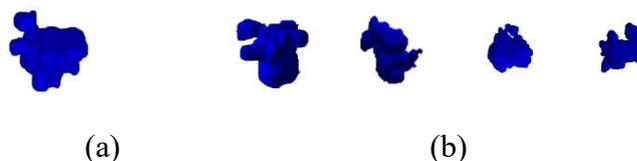
**Fig.12.** Classification results in confusion matrix

(a)



(b)

**Fig.13.** Examples of true positives and true negative (a) Nodules (b) Non-nodules



**Fig.14.** (a) Nodule wrongly classified as non-nodule (b) Non-nodules wrongly classified as nodules

Fig. 13 shows the examples of successfully classified nodules (true positives) and non-nodule (true negatives). Fig. 14 shows nodule which is wrongly classified as non-nodule (false negative) and non-nodules which are wrongly classified as nodules (false positives) of the classification. The wrong classification for false negative was occurred because of the attachment of the non-nodule candidates as shown in Fig. 14 (a). For the false positives in Fig. 14 (b), wrong classification was occurred because the non-nodule candidates appear in most of the layers of the 3D and it makes to increase the volume of the candidates.

For the computation time, the proposed system takes around 7 minutes for a case having 60 slices in average. The system is developed using MATLAB R2016b and the processor is Intel(R) Core(TM) i7-6500U CPU @ 2.60GHz, 8MB RAM. The processing time for each function in the proposed system is illustrated in Table 3.

**Table 3.** Processing time of the proposed system

Function Name	Calls	Total Time	Self-Time
ProposedCAD	1	422.706 s	0.820 s
FeaturesExtraction	8	412.655 s	412.079 s
Segmentation	65	6.048 s	0.287 s
imsegfmm	65	3.8837 s	0.055 s
fastmarchingmex (MEX-file)	65	3.722 s	3.722 s
RandomForestClassification	1	1.723 s	0.018 s

The comparison of the performance of the proposed method with others published 3D methods is described in Table 4.

#### 4. CONCLUSION

This research proposes a hybrid system to detect the pulmonary nodule using 2D nodule candidates segmentation, 3D nodule candidates reconstruction and feature extraction. Thirty eight (38) different cases of lung cancer from Lung CT-Diagnosis collection downloaded from TCIA are used to test the system. By applying the fast marching, the nodule candidates from

2D CT slices of each case are segmented first. Applying this segmentation method can ease the inconveniences of (i) appropriate threshold value selection in morphological and thresholding segmentation, (ii) noise sensitivity and over segmentation of watershed segmentation and (iii) time consuming and less sensitive to weak edge of active contour. The segmented nodule candidates from 2D slices are subsequently required for 3D reconstruction and feature extraction. The proposed hybrid method also can eliminate the computational effort of using only 3D. The random forest gives the classification accuracy of 92% with 4 false positives and 1 false negative. As a future work, the nodules outputted by the system can be used to classify benign and malignant nodules. Moreover, the system can be improved by extracting more 3D image features to get better results.

**Table 4.** Comparison of proposed system with literature

Methods	Samples	Segmentation	Classifier	Accuracy
[15]	144	Bounding box, Binarization, Rolling ball, Dot enhancement filter	SVM	86%
[16]	58	Active contour, 3D visualization technique	KNN	90%
[17]	30	3D multi-scale Block LBP filter	-	89.7%
[18]	65	Thresholding, 3D blob algorithm	SVM	89.19%
Proposed Method	61	Fast marching, 3D connected component labelling	Random forest	92%

## 5. ACKNOWLEDGEMENTS

Authors sincerely acknowledge to Asian University Network (AUN/SEED-Net) for financial support for this research. Moreover, we gratefully appreciate to TCIA for sharing lung CT scan images.

## 6. REFERENCES

- [1] World Health Organization (WHO). Cancer control: Knowledge into action: WHO guide for effective programmes. Geneva: WHO, 2007
- [2] U.S. Food and Drug Administration. Guidance for premarket notification [510(k)] submission for computer assisted detection devices. Maryland: United States Food and Drug Administration, 2012
- [3] American Cancer Society. Lung cancer prevention and early detection. 2016, <https://old.cancer.org/acs/groups/cid/documents/webcontent/acspc-039558-pdf.pdf>
- [4] Diagnostic Imaging Pathways. Solitary pulmonary nodule. 2012, <http://www.imagingpathways.health.wa.gov.au/index.php/imaging-pathways/respiratory/solitary-pulmonary-nodule>
- [5] Tariq A, Akram M U, Javed M Y. Lung nodule detection in CT images using neuro fuzzy classifier. In IEEE 4th International Workshop on Computational Intelligence in Medical Imaging, 2013, pp. 49-53
- [6] Mukherjee J, Chakrabarti A, Shaikh S H, Kar M. Automatic detection and classification of solitary pulmonary nodules from lung CT images. In IEEE 4th International Conference of Emerging Applications of Information Technology, 2014, pp. 294-299
- [7] Punithavathy K, Ramya M M, Poobal S. Analysis of statistical texture features for automatic lung cancer detection in PET/CT images. In IEEE International Conference on Robotics, Automation, Control and Embedded Systems, 2015, pp. 1-5
- [8] Nitheesha S, Sowmya B J. Lung cancer recognition in CT image using watershed segmentation. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 2016, 5(4):3405-3409
- [9] Kanitkar S S, Thombare N D, Lokhande S S. Detection of lung cancer using marker-controlled watershed transform. In IEEE International Conference on Pervasive Computing, 2015, pp. 1-6
- [10] Keshani M, Azimifar Z, Boostani R, Shakibafar A. Lung nodule segmentation using active contour modeling. In 6th IEEE Iranian Machine Vision and Image Processing, 2010, pp. 1-6
- [11] Roy T S, Sirohi N, Patle A. Classification of lung image and nodule detection using fuzzy inference system. In IEEE International Conference on Computing, Communication and Automation, 2015, pp. 1204-1207

- 
- [12] Nithila E E, Kumar S S. Segmentation of lung nodule in CT data using active contour model and Fuzzy C-mean clustering. *Alexandria Engineering Journal*, 2016, 55(3):2583-2588
- [13] Sifakis E, Tziritas G. Fast marching techniques for image segmentation. 2017, <http://www.csd.uoc.gr/~tziritas/papers/segmentSpringer.pdf>
- [14] Javaid M, Javid M, Rehman M Z, Shah S I. A novel approach to CAD system for the detection of lung nodules in CT images. *Computer Methods and Programs in Biomedicine*, 2016, 135:125-139
- [15] Liu Y, Yang J, Zhao D, Liu J. A study of pulmonary nodule detection in three-dimensional thoracic CT scans. In *2nd IEEE International Conference on Computer Modeling and Simulation*, 2010, pp. 481-484
- [16] Soltaninejad S, Keshani M, Tajeripour F. Lung nodule detection by KNN classifier and active contour modelling and 3D visualization. In *16th IEEE CSI International Symposium on Artificial Intelligence and Signal Processing*, 2012, pp. 440-445
- [17] Xu F, Zhang W J, Li X Y, Xiao H, Peng S H, Nam H D, Zhang M F. A 3D multi-scale Block LBP Filter for lung nodule enhancement based on the CT images. In *IEEE International Congress on Image and Signal Processing, BioMedical Engineering and Informatics*, 2016, pp. 1376-1380
- [18] Peña D M, Luo S, Abdelgader A. Auto diagnostics of lung nodules using minimal characteristics extraction technique. *Diagnostics*, 2016, 6(1):1-14
- [19] Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L. The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *Journal of Digital Imaging*, 2013, 26(6):1045-1057
- [20] Grove O, Berglund A E, Schabath M B, Aerts H J, Dekker A, Wang H, Velazquez E R, Lambin P, Gu Y, Balagurunathan Y, Eikman E. Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma. *Plos One*, 2015, 10(3):1-14
- [21] Sethian J A. A fast marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Sciences*, 1996, 93(4):1591-1595
- [22] Sethian J A. Theory, algorithms, and applications of level set methods for propagating interfaces. *Acta Numerica*, 1996, 5:309-395
- [23] Otsu N. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 1979, 9(1):62-66

- [24] Albert R H, Russell J J. Evaluation of the solitary pulmonary nodule. *American Family Physician*, 2009, 80(8):827-831
- [25] Duc M H, Peter J M. Pulmonary nodules. 2014, <http://www.clevelandclinicmeded.com/medicalpubs/diseasemanagement/hematology-oncology/pulmonary-nodules>
- [26] Guttentag A R. Thoracic fungal disease. 2017, <http://learningradiology.com/lectures/facultylectures/Thoracic%20Fungal%20Disease%20AG%20HTML5/Thoracic%20Fungal%20Disease/Thoracic%20Fungal%20Disease.html>
- [27] Ceylan N, Bayraktaroglu S, Savaş R, Alper H. CT findings of high-attenuation pulmonary abnormalities. *Insights Into Imaging*, 2010, 1(4):287-292
- [28] European Society of Radiology. Dr. Pepe's Diploma Casebook: Case 42-Solved! 2013, <http://blog.myesr.org/dr-pepes-diploma-casebook-case-42>
- [29] Ericsoussi B. Approach to the solitary pulmonary nodule new staging system for NSCLC lymph node map-update. 2010, <https://www.slideshare.net/basselericsoussi/approach-to-the-solitary-pulmonary-nodule-new-staging-system-for-nsclc-lymph-node-mapupdate>
- [30] Dillak R Y, Dana S, Beily M. Face recognition using 3D GLCM and Elman Levenberg recurrent Neural Network. In *IEEE International Seminar on Application for Technology of Information and Communication*, 2016, pp. 152-156
- [31] Haralick R M, Shanmugam K. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 1973, 6:610-621
- [32] Breiman L. Random forests. *Machine Learning*, 2001, 45(1):5-32

**How to cite this article:**

Paing MP, Choomchuay S. 3D pulmonary nodules detection using fast marching segmentation. *J. Fundam. Appl. Sci.*, 2017, 9(4S), 319-339.