Automatic Lung Nodule Detection from Chest CT Data Using Geometrical Features: Initial Results

Tarik A. Chowdhury, Paul F. Whelan
Centre for Image Processing and Analysis, Dublin City University, Dublin-9, Ireland
tarik.a.chowdhury@dcu.ie, paul.whelan@dcu.ie

Abstract

In this paper, a complete system for automatic lung nodule detection from Chest CT data is proposed. The proposed system includes the methods of lung segmentation and nodule detection from CT data. The algorithm for lung segmentation consists of surrounding air voxel removal, body fat/tissue identification, trachea detection, and pulmonary vessels segmentation. The nodule detection algorithm comprises of candidate surface generation, geometrical feature generation and classification. The proposed system shows 88.2% sensitivity for nodule >=3mm with 8.91 false positive per dataset.

1. Introduction

Lung cancer is the leading cause of cancer related death worldwide and approximately 1.3 million new cases are diagnosed every year [1]. Cancer statistics show that more people die of lung cancer than of colon, breast, and prostate cancers combined. The five years survival rate for all stages of lung cancer (surgery, radiation therapy, and chemotherapy treatment of lung carcinoma) combined is 14% [2]. The only prospect for curing early lung cancer is prompt diagnosis of Solitary Pulmonary Nodules (SPN) and subsequent treatment. CT is considered one of the most accurate medical imaging modalities for early detection of lung nodules [3]. A CT scan of a patient with reconstruction interval of 1mm offers 300-400 image slices for analysis by a radiologist for nodule detection. Examination of this volume of data is a time consuming and error prone task. This issue has motivated the development of methods for computer aided detection (CAD) of lung nodule from chest CT data. It is well accepted that CAD based automatic nodule detection can be used as a second reader with a radiologist for finding suspicious lesions in lung CT data [4].

The algorithm for lung nodule detection from chest CT data can be divided into four major steps. The first step is segmentation of lung from chest CT data. The second step is candidate generation from the segmented lung. The third step is feature generation and selection for the candidate surfaces. The fourth and final step is classification of candidate surfaces as nodule or non-nodule. In the last decade, a number of methods were reported for automatic segmentation of lung from chest CT data [5-9]. Brown et al. [5, 6] developed a method that employed a knowledge based anatomical model for lung segmentation. Their method [5,6] starts with image segmentation using gray level thresholding, 3-D region growing and mathematical morphology. Finally, a model based fuzzy set is employed to detect the lung from segmented area. Armato III et al. [8, 9] proposed a method that applied multiple gray level thresholding for lung segmentation. A rolling ball method was employed to segment the nodule in the lung wall. Hu et al. [7] developed a method that applied gray level thresholding for binary image creation. In the next step, dynamic programming was employed for detecting left and right lungs from the binary images. Finally, smoothing was applied at the lung boundary.

Regarding the existing techniques, a good survey for recent developments in nodule detection from chest CT data can be found in the report published by Qiang Li [10]. Mao et al. [11] applied a special circular filter named as 1.5-D circular pattern filter for nodule enhancement, where the filter was calculated using gradient information of the input image. The Bae et al. [12] CAD-scheme used four different morphological filters to enhance juxtapleural nodules. Bae et al. [12] also employed 2D and 3D elongation criteria derived from eigenvalues of the candidate volume for reduction of false positives (FPs). Overall sensitivity of the Bae et al. [12] developed CAD program was 95.1%. Gori et al. [13] presented a CAD scheme that included a method called dot-filter to enhance the nodule characteristic, and that applied a neural classifier to

reduce the false positives. Gori et al. [13] achieved a sensitivity of 85.3% at 6 FPs/scan. Li et al. [14] developed a CAD scheme that modified the enhancement filters developed by Koller et al. [15] to distinguish between nodules, vessels, and airway walls. These filters were based on the eigenvalue analysis of the Hessian matrix at each location in two-dimensional or three-dimensional image space. The Li et al. [14] selective nodule enhancement filter showed 93.47% (71/76) sensitivity with 4.2 FPs per section. Lee et al. [16] proposed a method based on a template-matching technique. The developed technique [16] employed genetic algorithm to locate the target position in the image using Gaussian template images as reference. Lee et al. [16] employed 557 sectional images from 20 clinical cases and 71/98 (approximately 72%) were correctly detected, with approximately 1.1 FPs/slice. Dehmeshki et al. [17] proposed a method that starts with lung segmentation and applies a spherical kernel for enhancement of lung parenchyma. In the next step of their method, shape based GATM were employed for candidate surface generation. For each candidate surface, six shape features were passed through a rule based classifier to declare the candidate as nodule or non-nodule. Bellotti et al. [18] devised a CAD scheme that employed region growing, active contours model, and convex hulls to generate candidate nodule surfaces. False positives were reduced by a two layered supervised neural network using geometrical feature of candidate surfaces as neuron. The detection rate of the system is 88.5% with 6.6 FPs/scan on 15 CT scans with 26 nodules. Other notable methods for lung nodule detection are reported in [19-21].

The aim of this paper is to detail the implementation of a CAD system that has been developed as a modular system where all components of nodule detection from chest CT data are interconnected. The proposed system initially detects the lungs from the CT data using the methods of surrounding air voxel removal, body tissue/fat/bone identification, trachea detection and iterative labelling. In the next step, nodules are detected from the segmented lungs by employing the methods of surface normal concentration, histogram based candidate generation, morphological features and nearest neighbourhood classification.

2. Lung Nodule Detection

The proposed method is designed for detection of lung nodules that are within the size range of 3-30mm only. The proposed system can be divided into four main technical steps. The steps are lung segmentation, candidate surface generation, feature extraction and classification (see Figure 1).

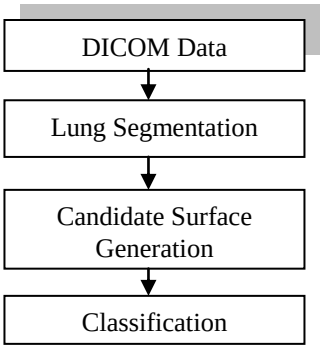


Figure 1: Flow chart for nodule detection from Chest CT data.

2.1. Lung Segmentation

The method of lung segmentation comprises of surrounding air voxel detection, fat and body tissue identification, lung area detection and pulmonary vessel extraction. Surrounding air voxel removal starts with applying a 2D region growing algorithm in each slice of the CT data. The four corners of each slice are considered as seed points for 2D region growing with a threshold of -700 Hounsfield unit (HU). After removal of surrounding air voxels, remaining voxels in the patient data are body fat, bone, body tissue, blood vessels and air voxels in the lungs. Any voxel that is neighbour to tagged surrounding air voxels with HU value higher than -700HU is considered as seed point for 2D region growing to identify the body fat/tissue/bone. The next step of lung segmentation is trachea detection. In general, the trachea can be found within the first few slices of the patient data. The shape of the trachea is similar to a tube, hence a 2D view of the trachea in an axial CT scan looks similar to a circle. If the labelling fails to detect an air region in the first slice of the patient data, the method is re-applied in the next slice. This labelling process continues until the method finds an air area larger than a threshold size. Next, we apply circle fitting on the edge of the labelled area. If the circle fitting error is less than a threshold and the circle centre passes a neighbourhood test, we consider the labelled area as part of the trachea. If multiple air areas are found, the area with the best circle fit is selected as a candidate for the trachea. The centre of the detected trachea area is considered as a seed point for 3D region growing to detect lung from CT data. The process of 3D region growing is applied iteratively while changing the threshold from-990HU to -700HU. At the end of the 3D growing processes, all

air voxels inside the lungs are detected and are marked as lung voxels. Remaining voxels with HU value higher than -700 inside the lungs area are pulmonary vessels/blood vessels.

2.2. Candidate Surface Generation

The algorithm for candidate surface generation consists of the steps of derivation of surface normal, generation of candidate centre, creation of histogram based candidate surface and performing of convexity test (see Figure 2).

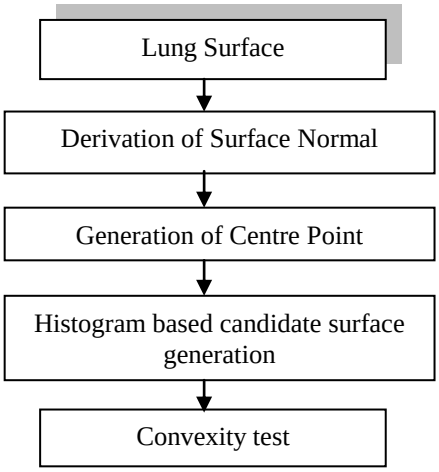


Figure 2: Algorithm for candidate surface generation

Candidate surface generation starts with surface normal calculation. The 3D Zuker-Hummel operator was employed to derive the normal of each surface voxel. The generation of candidate centre starts with finding seven 3D points in the direction of the surface normal. 3D points are evenly distributed from 1.5mm to 8.50mm. Initially, a histogram of these 3D points is calculated. Smoothing is applied to reduce the effect of noise in the direction of the surface normal [22]. Non-maximum suppression is employed to generate the candidate centres [22]. A Gaussian distribution is applied to find the refined candidate centres [22]. Now for each candidate centre, a distance histogram based candidate surface generation method [23] is applied to derive the candidate surfaces. In this method, the distance histogram is created using the Euclidian distances between the candidate centre and the sub-volume surface voxels ($40x40x40mm^3$ cubic volume). Figure 3 shows a typical distance histogram for a candidate (blue line/upper line). The red line/lower line in Figure 3 shows the first derivative of the histogram. Let d_f and d_s be the first and second zero crossing distances from the candidate centre (see Figure 3, marked with two vertical lines). Then $d_{mean} = (d_f+d_s)/2$ is considered as threshold distance for candidate surface generation. All surface voxels of the sub-volume within the distance of d_{mean} from the candidate centre are clustered as candidate surface. Finally, we apply a surface normal intersection [22] method to remove non-convex surface voxels from the candidate cluster (see Figure 3).

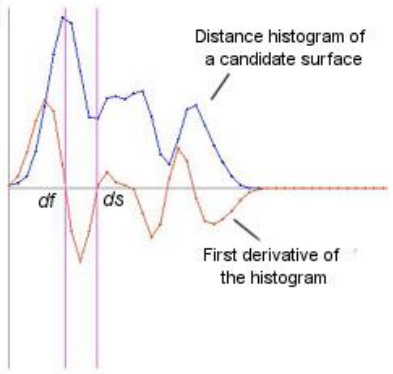
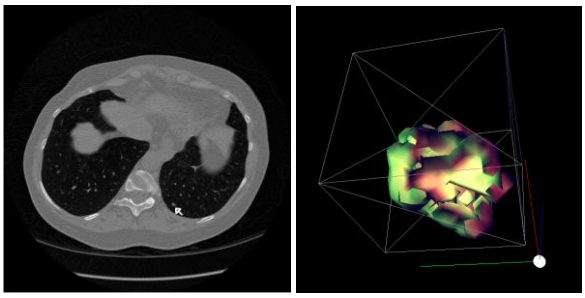


Figure 3: Candidate surface generation. Distance histogram - The original distance plot is shown in blue (upper line), indicating the number of surface voxels in the subregion at given distances from the initial seed point. The plot in red (lower line) indicates the first derivative of the histogram, where the vertical lines mark the maximum/minimum pair in the plot that correspond to the range of surface voxels selected for analysis.



(a) (b)

Figure 4: Candidate surface of a nodule. (a) is 2D axial view of a nodule. (c) is the 3D surface of the nodule.

2.3. Feature Generation

The next technical step of nodule detection is feature extraction. It is useful to note that nodules in the pulmonary vessels are spherical/ellipsoidal in shape and nodules in the lung wall are half spherical/ellipsoid in shape. Using the knowledge that the nominal model for nodules is spherical/ellipsoidal and the nominal model for pulmonary vessels is cylindrical, we extracted a number of key features, namely: standard deviation (SD) of three axes of ellipsoid and of sphere

radius, SD of surface variation, sphere radius, principal axis of ellipsoid fitting, Gaussian distribution value of the centre [22].

2.4. Classification

The fourth and final technical step of nodule detection is the classification of candidates as either nodules or non-nodules. In this regard, we applied a feature normalized nearest neighbourhood (FNNN) classifier [22]. The FNNN classification scheme consists of two standard stages. Firstly, the training database is created using the features detailed in the previous section for each class of pulmonary nodules, lung wall nodules and pulmonary vessels. The features of each class were normalised in order to avoid the situations where the features with largest values overpower the remaining ones [22]. The pulmonary nodule database uses three classes of nodules: small (32 nodules), medium (21 nodules), and large (26 nodules), and five classes of pulmonary vessels: large (117), medium (66), small (127), small convex surface (67) and cylindrical (103). The lung wall database contains three classes of nodules: large (9), medium (20), and small (23), and two classes of lung wall non-nodule candidate: large (127) and small (41).

3. Results and Discussion

The patient dataset are acquired from the National Biomedical Imaging Archive [24]. The Archive contains LIDC (Lung Imaging Database Consortium) and RIDER datasets on Chest CT. Each of the patient datasets in LIDC contains an annotation file which is in XML file format. In the XML file readers, (Radiologists) marks are stored in three categories. The first category consists of the nodule information for nodules of size 3mm and larger. In this category, each reader drew a complete outline around the nodule in all sections in which the nodule appears, and the X,Y,Z coordinates of nodule surfaces were stored in the XML file. The second category consists of the nodule information on nodules of size less than 3mm. In this scenario, readers indicate only the approximate centre of mass for each nodule (X,Y, Z coordinate). The third category describes the non-nodules that are 3mm and larger in diameter. Readers indicate only the centre of the mass of these non-nodules. Each patient data is reviewed by at least two radiologists and the XML file contains reports from each radiologist. The reconstruction interval (RI) for the patient data varies within the range 0.625mm-3mm (RI: 0.625mm - 3patients; 1mm - 3patients; 1.25mm - 30patients; 1.8mm -30patients; 2.0mm - 6patients; 2.5mm – 24patients and 3.0mm – 14 patients). The tube current of the patient datasets is in the range of 40mA to 200mA. The method of lung segmentation was applied to 110 LIDC patient datasets. The method failed to detect Trachea in one patient data and consequently failed to segment the lungs. Overall success rate of lung segmentation is 99%.

The weakness of the proposed lung segmentation method is the predefined threshold selection (-700HU) in 3D region growing. Patient movement and low dose data acquisition can create large artefact in the CT data and results in erroneous candidate surfaces generation. This error in candidate surface generation creates false positive and true negative classification in nodule detection. The development of an automatic threshold selection method based on noise analysis of the CT data would be desirable and is considered as future work for lung segmentation.

We have applied the proposed method on 95 patient datasets for nodule detection. The remaining 14 (109) patient datasets are employed for database creation for the classifier. The 95 patient datasets include 485 nodules in which 274 nodules are less than 3mm (< 3mm), and 211 nodules are 3mm and larger (>= 3mm). The overall sensitivity of nodule detection is 48.7%. The sensitivity of nodule detection for 3mm and larger (>= 3mm) is 88.2% (see Table 1). The proposed method missed 25 nodules which are clinically significant (>= 3mm). Of the 25 missed nodules 9 are 5mm and larger. The proposed method does not include the methodology for detection of non-solid nodules. We employed the features that are calculated based on the assumption of the spherical/ellipsoidal nature of nodules. A nodule can be non-solid, partly non-solid, and can be completely different from the spherical/ellipsoidal shape. If the candidate is irregular in shape or is too large (mass), the method of candidate surface generation detects the nodule/mass only partially. Hence, geometrical features for a partial candidate surface of a mass/spiculate nodule are different to the features generated from a nodule surface and classifier detects this candidate as non-nodule. Further work is being carried out to detect non-solid nodules, masses and to reduce false positives.

Table 1: Results of automatic computer aided nodule detection

Nodule Type	No.	TP	Sensitivity (%)
<3mm	274	50	18.2
>=3mm	211	186	88.2
False Positive			8.91 per patient

The proposed method generates on average 1221 candidate surfaces per patient. After applying the nearest neighbourhood classifier, the current method detects 846 false positive in 95 patient datasets (8.91 false positive per patient). In our experiments, we use the LIDC datasets that are scanned with 0.625mm - 3.0mm reconstruction interval. Only six datasets had reconstruction interval were less than or equal to 1.0mm. Hence, in 89 datasets nodules with diameter less than 2mm are seen in one-two slice and the current method failed to detect the small nodules (< 2mm).

Table 2: Performance levels for various techniques reported in [5-21].

Authors	Patient Numbers	Size (mm)	Sensitivity (%)	False Positive
Brown et al. [5]	29	<3 >=3	70 100	15pp
Bae et al, [12]	20	3 – 5 >5	91.2 97.2	6.9 pp 4.0 pp
Li et al [14]	32	8-13 >13	94% 95%	6.6 pp
Dehmeshki et al. [17]	70	3-20	90%	14.6pp
Bellotti et al [18]	15	5-15	88.5%	6.6pp
Zhang et al [19],	17	2 – 15	85.6%	9.5pp
Paik et al. [20]	8	>=6	80 90 95	1.3pp 5.6pp 63pp
Ge et al [21]	56	3-30	80	0.37 per section

4. Conclusion

Our CAD system shows high sensitivity for detecting nodules larger than 3mm while maintaining a low level of false positive per patient. Experimental results show that our method is comparable to other methods reported [5-21] for detecting nodules larger or equal to 3mm. Results of the current system indicate that the system can be applied for clinical studies.

5. Acknowledgments

The authors would like to acknowledge the valuable input from our colleagues in the Centre for Image Processing & Analysis, namely, Robert Sadleir and Ovidiu Ghita. This work was funded by the Rince Institute (DCU), and the National Biophotonics and Imaging Platform Ireland [NBIP] (HEA-PRTLI IV).

6. References

[1] J. Ferlay, P. Autier, M. Boniol, M. Heanue, M. Colombet and P. Boyle, "Estimates of the cancer incidence and mortality in Europe in 2006", *Annals of Oncology*, 2007, 18(3): 581-592.

[2] M.J. Horner, L.A.G. Ries, M. Krapcho, N. Neyman, R. Aminou, N. Howlader, S.F. Altekruse, E.J. Feuer, L. Huang, A. Mariotto, B.A. Miller, D.R. Lewis, M.P. Eisner, D.G. Stinchcomb, B.K. Edwards, "SEER Cancer Statistics Review, 1975-2006", *National Cancer Institute*, Bethesda, MD, 2009.

[3] S. Sone, F. Li, Z.G. Yang, S. Takashima, Y. Maruyama, M. Hasegawa, J.C. Wang, S. Kawakami and T. Honda, "Characteristics of small lung cancers invisible on conventional chest radiography and detected by population based screening using spiral CT", *The British Journal of Radiology*, 2000, 73(866); 137-145.

[4] S.G. Armato, F. Li, M.L. Giger, H. MacMahon, S. Sone, K. Doi, "Lung cancer: performance of automated lung nodule detection applied to cancers missed in a CT screening program", *Radiology*, 2002, 225:685-92.

[5] M.S. Brown, M.F. McNitt-Gray, N.J. Mankovich, et al, "Method for segmenting chest CT image data using an anatomical model: preliminary results", *IEEE Transaction on Medical Imaging*, 1997, 16:828-839.

[6] M.S. Brown, M.F. McNitt-Gray, J.G. Goldin, D.R. Aberle, "Extensible knowledge-based architecture for segmenting CT data", *In: Hanson KM, ed. Medical imaging 1998: image processing. Proc SPIE*, 1998, 3338: 564-574.

[7] S. Hu, E.A. Hoffman, J.M. Reinhardt, "Automatic Lung Segmentation for Accurate Quantitation of Volumetric X-Ray CT Images", *IEEE Transaction on Medical Imaging*, June 2001, 20, 6.

[8] S.G. Armato III, W.F. Sensakovic, "Automated Lung Segmentation for Thoracic CT: Impact on Computer-Aided Diagnosis", *Acad Radiol*, 2004, 11:1011-1021.

[9] S.G. Armato, H. MacMahon, "Automated lung segmentation and computer-aided diagnosis for thoracic CT scans", International Congress Series, 2003, 1256, 977- 982.

[10] Q. Li, "Recent progress in computer-aided diagnosis of lung nodules on thin-section CT", *Computerized Medical Imaging and Graphics*, 2007, 31 248-257.

[11] F. Mao, W. Qian, L.P. Clarke, "1.5 Dimensional Circular Pattern Filter for Multi-scale Lung Nodule Detection", *18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 31 Oct. 3 Nov. 1996.

[12] K.T. Bae, J.S. Kim, Y.H. Na, K.G. Kim, J.H. Kim, "Pulmonary nodules: automated detection on CT images with morphologic matching algorithm - preliminary results", *Radiology*, 2005, 236, 286-294.

[13] M.E.F. Gori, M.E. Fantacci, A.P. Martinez, A. Retico, "An automated system for lung nodule detection in low-dose computed tomography", *Proc. SPIE*, 2007, 6514, 65143R.

[14] Q. Li, S. Sone, K. Doi, "Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans", *Medical Physics* 2003, 30:2040-2051.

[15] T.M. Koller, G. Gerig, G. Zekely, and D. Dettwiler, "Multiscale detection of curvilinear structures in 2-D and 3-D image data", *International Conference on Computer Vision*, IEEE Computer Society Press, Los Alamitos, 1995, 864-869.
[16] Y. Lee, T. Hara, H. Fujita, S. Itoh, T. Ishigaki, "Automated detection of pulmonary nodules in helical CT images based on an improved template matching technique", *IEEE Transactions on Medical Imaging* 2001, 20, 595-604.
[17] J. Dehmeshki, X. Ye, X. Lin, M. Valdivieso, H. Amin, "Automated detection of lung nodules in CT images using shape-based genetic algorithm", *Computerized Medical Imaging and Graphics*, 1007, 31 (6), 408-417.
[18] R. Bellotti, F. De Carlo, G. Gargano, S. Tangaro, D. Cascio, E. Catanzariti, P. Cerello, S.C. Cheran, P. Delogu, I. De Mitri, C. Fulcheri, D. Grosso, A. Retico, S. Squarcia, E. Tommasi, B. Golosio, "A CAD system for nodule detection in lowdose lung CTs based on region growing and a new active contour model", *Medical Physics*, 2007, 34 (12), 4901-4910.
[19] X. Zhang, G. McLennan, E.A. Hoffman, M. Sonka, "Automated detection of small-size pulmonary nodules based on helical CT images", *Information Processing in Medical Imaging*, 2005, 3565, 664-676.
[20] D.S. Paik, C.F. Beaulieu, G.D. Rubin, B. Acar, R.B. Jeffrey, Jr., J. Yee, J. Dey, and S. Napel, "Surface normal overlap: a computer-aided detection algorithm with application to colonic polyps and lung nodules in helical CT", *IEEE Transactions on Medical Imaging*, 2004, 23 (6), 661-675.
[21] Z.Y. Ge, B. Sahiner, H.P. Chan, L.M. Hadjiiski, P.N. Cascade, Bogot N, Kazerooni, EA, Wei, J, Zhou, C, "Computer-aided detection of lung nodules: false positive reduction using a 3D gradient field method and 3D ellipsoid fitting", *Medical Physics*, 2005, 32 (8), 2443-2454.
[22] T.A. Chowdhury, P.F. Whelan and O. Ghita, "A Fully Automatic CAD CTC System Based on Curvature Analysis for Standard and Low Dose CT Data", *IEEE Transactions on Biomedical Engineering*, March 2008, 55(3), 888-901.
[23] N. Sezille, P.F. Whelan, Kevin Robinson, "A method for detecting at polyps in the colon", EU Patent Filed Application No: 07121814.3 & US Patent Application No: 60/990,725, 2007.
[24] National Biomedical Imaging Archive (https://imaging.nci.nih.gov/ncia/login.jsf).
http://www.cancer.gov/.