ISTANBUL UNIVERSITY –
JOURNAL OF ELECTRICAL & ELECTRONICS ENGINEERING

YEAR : 2006 (61 - 67)
VOLUME : 6
NUMBER : 1

RULE BASED DETECTION OF LUNG NODULES IN CT IMAGES

Serhat ÖZEKES [1] A.Yılmaz ÇAMURCU [2]

[1] Istanbul Commerce University, Ragip Gumuspala Cad. No: 84 Eminonu 34378 Istanbul, Turkey,

[2] Marmara University, Goztepe, 81040, Istanbul, Turkey

[1]Email: serhat@iticu.edu.tr [2]Email: camurcu@marmara.edu.tr

ABSTRACT

In this paper, we present a computer aided diagnosis (CAD) system for lung nodule detection in computed tomography (CT) images. Here, the density values of pixels in CT image slices are used and scanning the pixels in 8 directions is evaluated. By using various thresholds while scanning the pixels, lung nodule shapes and parts of the normal structure shapes (blood vessels, bronchus etc.) are found. All shapes are labeled using connected component labeling (CCL). Two rules are used to distinguish lung nodules from normal structures. In the first rule, the euclidean distance of the shape, and in the second rule the regularity which is the ratio of euclidean distance to thickness of the shape is considered. The performance of the system is evaluated using a test set which contains totally 35 normal and abnormal images, with 61 nodules. When results are compared with the second look reviews of a chest radiologist, it is seen that the system achieved 89% sensitivity with 0.457 false positives (FPs) per image. The proposed system which obtains high sensitivity with acceptable low number of false positives per image, may improve the computerized analysis of lung CTs and early diagnosis of lung nodules.

Keywords: *Computer aided diagnosis, lung nodule detection, computed tomography*

1. INTRODUCTION

According to the World Health Organization (WHO), lung cancer is one of the world's most common cancers, accounting for more than 1.2 million new cases annually. Lung cancer, often presented in lung nodules, can be difficult to detect in its early stages. A pulmonary nodule is generally defined as a rounded opacity, at least moderately well marginated and no greater than 3 cm in maximum diameter [1]. The most common causes for nodules detected by chest radiograph are granulomatous disease and lung cancer [2]. Other etiologies include solitary pulmonary metastases, hamartomas, and carcinod tumors [3,4]. Approximately 30% to 40% of solitary pulmonary nodules identified by chest radiography are malignant [3,4]. The American Cancer Society (ACS) reports that many early stage lung cancers are diagnosed incidentally as a result of imaging studies requested for an unrelated medical condition. The discovery of X-rays by Wilhelm Conrad Röntgen, in 1895 [5], has revolutionized the field of diagnostic medicine [6]. Today many types of examinations have been replaced by computed tomography (CT) and/or magnetic resonance imaging (MRI). However, the accuracy and efficiency of viewing hundreds of source axial

Received Date : 26.11.2004
Accepted Date: 26.10.2005

images per exam are limited by human factors, such as attention span and eye fatigue.

Researchers have developed various computerized lung nodule detection methods. For detecting lung nodules, Giger *et al.* [7] have used multiple gray-level thresholds to extract candidates, then, calculated 2-D geometric features such as perimeter, compactness and circularity for each candidate at each threshold. A rule-based approach was used to assign a confidence rating, in the range 1–5, to each 2-D candidate, with 1 being definite vessel and 5 being definite nodule. Confidence ratings were modified based on ratings of the nodule in adjacent slices. 94% per-nodule sensitivity was achieved with 1.25 false positives (FPs) per patient. Armato *et al.* [8], [9] applied multilevel thresholding and a rolling ball algorithm toward detecting lung nodules. Shape and attenuation features were classified using linear discriminant analysis and the algorithm achieved 70% per-nodule sensitivity with 1.5 FPs per axial section. Brown *et al.* [10] have presented an algorithm for both detection and surveillance of lung nodules in CT. Region-growing and morphological operators were used to create candidate locations. Attenuation, location, volume, and shape features were matched to model objects in a semantic net with fuzzy membership that serves as a generic a priori anatomic model. In the initial detection task, 86% per-nodule sensitivity was achieved with 11 FPs per patient.

In this study, we design a CAD system for lung nodule detection in CT images. Using the density values of pixels in image slices and scanning these pixels in 8 directions with distance thresholds, lung nodule shapes and parts of the normal structure shapes (blood vessels and bronchus) are found. All shapes are labeled using CCL. Then, two rules are used to distinguish lung nodules from normal structures. Hence, the true lung nodules are detected successfully.

2. MATERIALS AND METHODS

Each CT slice used in this study is formed of a 16 bit ASCII coded matrix with dimensions 512 X 512. These ASCII codes are related with the density values of each pixel in slices. Using these ASCII codes density values in Hounsfield units (HU) are calculated. HU is a unit of x-ray attenuation used for CT scans, each pixel being assigned a value on a scale on which air is -1000, water is 0, and compact bone is +1000 [11]. The dataset used in this study is formed of training set and test set. While system is designed using the training set, test set is used to evaluate the efficiency of the proposed system. For the training set 14 abnormal slices with 43 nodules and 6 normal slices are chosen randomly. One of these slices can be seen in Figure 1. When this training set is examined, it is determined that density values of the nodules are between *"minimum density threshold"* and *"maximum density threshold"* values.

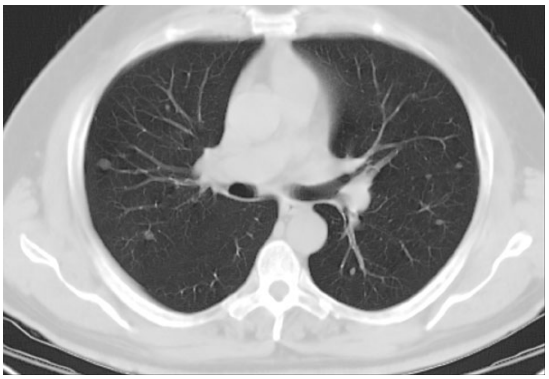

Figure 1. A CT image slice

2.1 EXTRACTION OF CANDIDATE LUNG NODULE REGION

Pixels, which form the candidate lung nodule region, must be members of a set of adjacent neighbour pixels with densities between *"minimum density threshold"* and *"maximum density threshold"* values. It has been observed that diameters of lung nodules are between upper and lower boundaries. So, to understand whether a pixel is in the center region of the shape, first, diameter of the shape (assuming the pixel in question is the center) should be considered. In this stage, we introduce two thresholds which form the boundaries. As seen in Figure 2, one is the "*minimum distance threshold*" representing the lower boundary and the other is the "*maximum distance threshold*" representing the upper boundary.

If a pixel has adjacent neighbours that are less than "*minimum distance threshold*" or more than "*maximum distance threshold*" in 8 directions, it could be concluded that this pixel couldn't be a part of candidate lung nodule. Otherwise, it could be a part of candidate lung nodule. Examples of determining the pixels to be a part of candidate lung nodule can be seen in Figure 3.

While the pixel under investigation in Figure 3a and 3b is not a part of the candidate lung nodule, in figure 3c it is a part of the candidate lung nodule. The values of minimum and maximum distance thresholds are dealt with the resolution of the CT image. These thresholds are used to avoid very big or very small structures such as parts of chest bones or heart and vertical vessels.

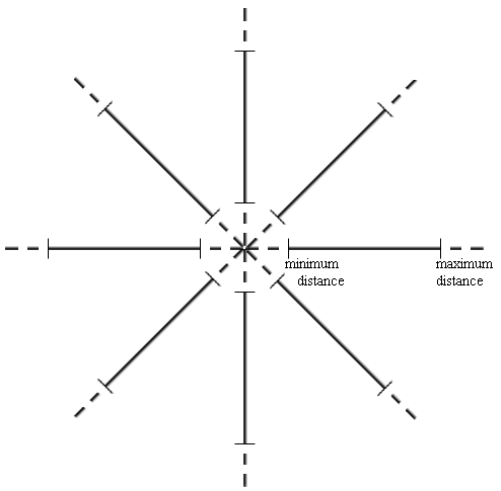

Figure 2. Minimum and maximum distance thresholds in 8 directions

At first step, to find all shapes (including lung nodules, blood vessels and other structures) which have density values between "*minimum density threshold"* and "*maximum density threshold"*, 0 pixel is assigned to "*minimum distance threshold*" and 5 pixels are assigned to "*maximum distance threshold*". Assigning 0 to "*minimum distance threshold*", we aimed to find all the shapes conforming the density constraints. The reason of choosing 5 as "*maximum distance threshold*" is that, it is the best and the most efficient number to be assigned which have been found by trial and error method. If the pixel meets conditions explained above, pixel's coordinates are recorded to be one of the pixels forming the candidate lung nodule region of the slice. As seen in Figure 4 in grey color, at the first step scanning the CT by assigning 0 pixel to "*minimum distance threshold*", shapes including blood vessels, lung nodules and other structures which have suitable density values, are found.

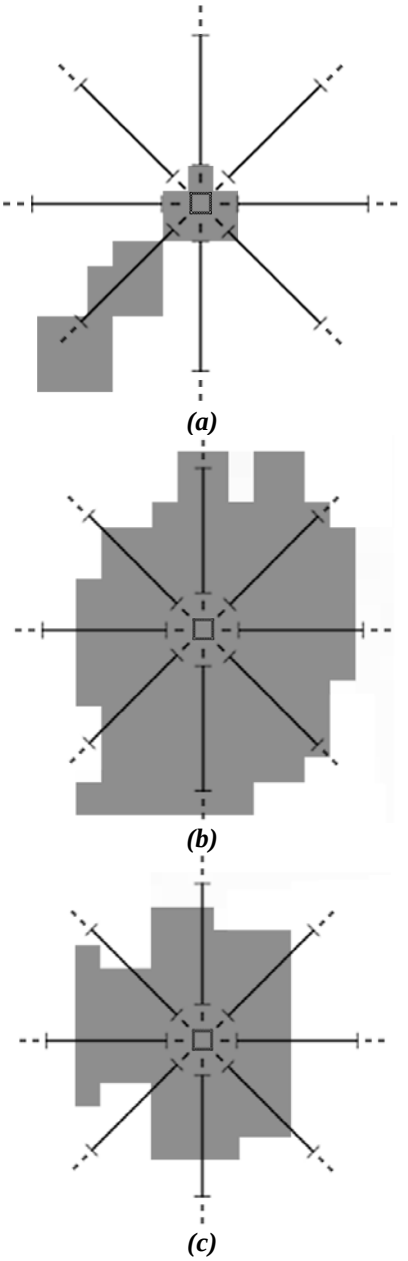

Figure 3. Examples of scanning with minimum and maximum distance thresholds

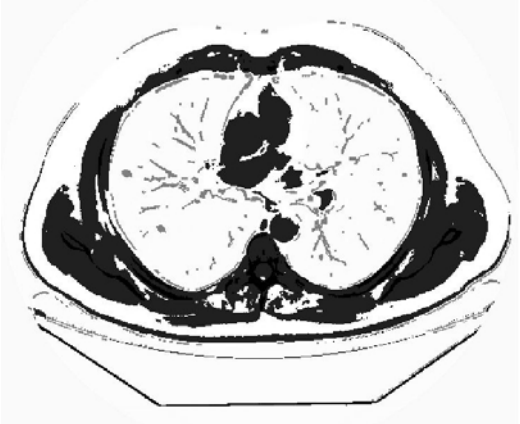

Figure 4. Lung nodules, blood vessels and other structures with suitable density values when "minimum distance threshold" is 0 pixel and "maximum distance threshold" is 5 pixels

At second step, to eliminate small and thin shapes, 1 pixel is assigned to "*minimum distance threshold*" and 5 pixels are assigned to "*maximum distance threshold*". If the pixel meets the conditions above, pixel's coordinates are recorded to be one of the pixels forming the candidate lung nodule region of the slice. As seen in Figure 5, scanning the CT by assigning 1 pixel to "*minimum distance threshold*", we find shapes which are more similar to some parts of blood vessels and candidate nodules. If a shape is too small and thin, its pixels couldn't have a number of adjacent neighbour pixels greater than or equal to the value of "*minimum distance threshold*". So, it should be erased.

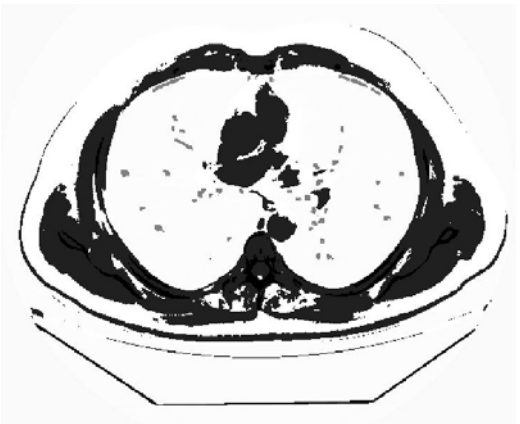

Figure 5. Lung nodules and parts of blood vessels remained with suitable density values when "minimum distance threshold" is 1 pixel and "maximum distance threshold" is 5 pixels

2.2 RULE BASED DETECTION OF TRUE LUNG NODULES

The problem here is that, some shapes look like nodules as in Figure 5, which are actually parts of the blood vessels. At the third step, to find the real lung nodules by distinguishing them from normal structures such as blood vessels and bronchus, we look at Figure 4, label each shape by using connected components labeling (CCL) and record the coordinates of each pixel of each shape.

CCL works by scanning an image, pixel by pixel (from top to bottom and left to right) in order to identify connected pixel regions - *i.e.* regions of adjacent pixels which share the same set of intensity values *V* [12]. The CCL operator scans the image by moving along a row until it comes to a point *p* (where *p* denotes the pixel to be labeled at any stage in the scanning process) for which *V={1}*. When this is true, it examines the four of the neighbours of *p* which have already been encountered in the scan (*i.e.* the neighbours (i) to the left of *p*, (ii) above it, and (iii and iv) the two upper diagonal terms). Based on this information, the labeling of *p* occurs as follows [13,14]:

- if all four neighbours are 0, assign a new label to *p*, else
- if only one neighbour has *V={1}*, assign its label to *p*, else
- if one or more of the neighbours have *V={1}*, assign one of the labels to *p* and make a note of the equivalences.

After completing the scan, the equivalent label pairs are sorted into equivalence classes and a unique label is assigned to each class. As a final step, a second scan is made through the image, during which each label is replaced by the label assigned to its equivalence classes. For display, the labels might be different grey levels or colors.

As seen in Figure 4, grey colored shapes are lung nodules, blood vessels and other structures which have densities between "*minimum density threshold"* and *"maximum density threshold"*. Also by looking at Figure 4, it can be realized that blood vessels and lung nodules have different morphologies. While lung nodules are thicker and circular, blood vessels are thinner and longer. So, to distinguish true lung nodules from normal lung structures (mainly blood

vessels) by using their morphologies, 2 rules are taken into account. In the first rule, the euclidean distance of the shape and in the second rule the regularity which is the ratio of euclidean distance to thickness of the shape is considered. We now explain these rules explicitly.

Rule 1: As mentioned above, lung nodules are circular, but blood vessels are long and thin. We can distinguish these two structures by calculating their euclidean distances. Because blood vessels are longer than lung nodules, it is expected that their euclidean distances are bigger. So, we use an "*euclidean distance threshold*" and assume that, for a shape to be a blood vessel candidate, its euclidean distance must be bigger than the "*euclidean distance threshold*". Otherwise, it would be a lung nodule candidate. The euclidean distance is calculated as follows:

$$E_d = \sqrt{(x_{\max} - x_{\min})^2 + (y_{\max} - y_{\min})^2} \quad (1)$$

Rule 2: As mentioned above, blood vessels are thinner than lung nodules. We can distinguish these two structures by calculating their thickness and regularity. The regularity of a shape is the ratio of its euclidean distance to its thickness. Because blood vessels are thinner than lung nodules, it is expected that their thickness is smaller and their regularity is bigger. So, we look for the thickness to compare the regularity of the shape with "*regularity threshold*". For a shape to be a candidate blood vessel, its regularity must be bigger than the "*regularity threshold*". Otherwise, it would be a lung nodule candidate.

To calculate the thickness of the shape, we first find the center coordinates. The center coordinates (x_c,y_c) are calculated as follows:

$$x_c = fix\left(x_{\min} + \frac{x_{\max} - x_{\min}}{2}\right) \quad (2)$$

$$y_c = fix\left(y_{\min} + \frac{y_{\max} - y_{\min}}{2}\right) \quad (3)$$

As a further step, $y_{c\,\min}$ (minimum y coordinate corresponding to x_c), $y_{c\,\max}$ (maximum y coordinate corresponding to x_c), $x_{c\,\min}$ (minimum x coordinate corresponding to y_c) and $x_{c\,\max}$ (maximum x coordinate corresponding to y_c) coordinates are found. Using these values, the height and width are calculated as follows:

$$height = y_{c\max} - y_{c\min} \quad (4)$$

$$width = x_{c\max} - x_{c\min} \quad (5)$$

To find the minimum thickness of the shape which we denote by *S*, the smaller one of height and width is chosen for the reasons explained above:

$$S = \min(height, width) \quad (6)$$

And finally, the regularity (*R*) is calculated as follows:

$$R = \frac{E_d}{S} \quad (7)$$

If these 2 rules are both valid for the same shape, it means that the euclidean distance of the shape is bigger than the "*euclidean distance threshold*" and the regularity of the shape is bigger than the "*regularity threshold*", the shape is a blood vessel. Because shape is not a true lung nodule, it is erased from Figure 5. These operations are repeated for all shapes and finally Figure 6 is obtained showing only true lung nodules with grey color.

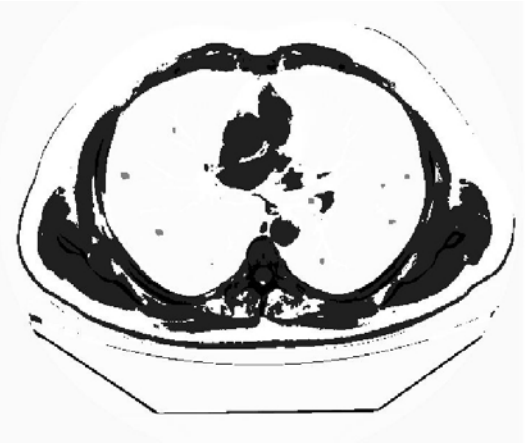

Figure 6. True lung nodules in the CT image slice.

3. RESULTS

The efficiency of our CAD system evaluated using the test set which contains 22 abnormal images with 61 nodules and 13 normal images. For various "*euclidean distance threshold*" and "*regularity threshold*" values the following results are achieved: 0.76 sensitivity with 0.142 FP/image, 0.84 sensitivity with 0.257 FP/image,

0.89 sensitivity with 0.457 FP/image, 0.93 sensitivity with 0.828 FP/image and 0.95 sensitivity with 1.4 FP/image.

Figure 7 shows the free-response receiver operating characteristic (FROC) curve which is a plot of operating points showing the tradeoff between the sensitivity versus the average number of false positives per image. See Figure 8 for the detected nodules and a FP in a slice. Nodules are shown with white circles and the FP is shown with a black circle.

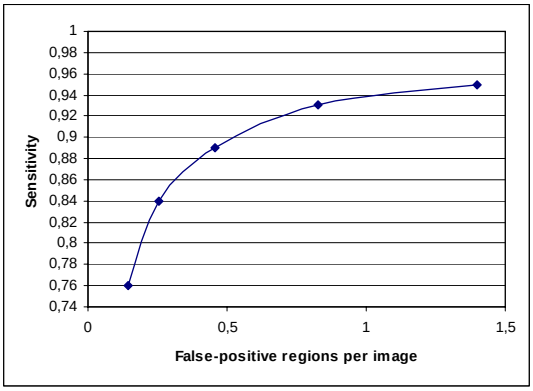

Figure 7. Free-response receiver operating characteristic curve showing the performance of the nodule detection task

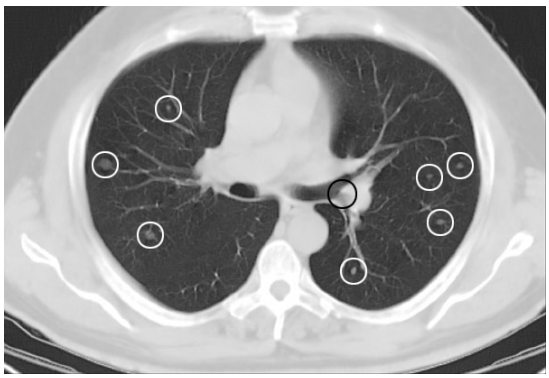

Figure 8. The detected true lung nodules in a CT image slice

4. CONCLUSIONS

In this paper, we develop a CAD system to automatically detect lung nodules in CT images. Use of CAD as a second reader may not only decrease the number of missed nodules, but also improve clinical efficiency. Morphological analysis of nodules is facilitated by computerized techniques. With computer-aided analytical techniques, one is more capable of studying the internal architecture of nodules.

This diagnostic system is constructed on image processing techniques and medical knowledge. We present the effectiveness of our system by applying it to patient's data and comparing the results with the second look reviews of a chest radiologist. The results show that when fully developed, a CAD system may improve the accuracy and efficiency in lung nodule detection.

REFERENCES

[1] Austin J. H., Muller N. L., Friedman P. J., vd. "Glossary of terms for CT of the lungs: Recommendations of the Nomenclature Committee of the Fleischner Society", *Radiology*, 200: 327-331, 1996.

[2] Keagy B. A., Starek P. J., Murray G. F., vd. "Major pulmonary resection for suspected but unconfirmed malignancy", *Ann Thorac Surg*, 38, 314-316, 1984.

[3] Siegeman S. S., Khouri N. F., Leo F. P., vd. "Solitary pulmonary nodules: CT assessment", *Radiology*, 160, 307-312, 1986.

[4] Ko J. P., Naidich D. P., "Lung nodule detection and characterization with multislice CT", *Radiologic Clinics of North America*, 41, 575-597, 2003.

[5] Röntgen W., "Über eine neue art von strahlen", in Sitzungsberichte der Physikalisch-Medicinisch Gesellschaft zu Würzburg, pp. 132–141, 1895.

[6] Paik D. S., Beaulieu C. F., Rubin G. D., Acar B., Jeffrey R. B., Yee Jr., J., Dey J., and Napel S., "Surface Normal Overlap: A Computer-Aided Detection Algorithm With Application to Colonic Polyps and Lung Nodules in Helical CT", *IEEE Trans. Med. Imag.*, vol. 23, no. 6, June 2004.

[7] Giger M. L., Bae K. T., and MacMahon H., "Computerized detection of pulmonary nodules in computed tomography images," *Investigat. Radiol.*, vol. 29, pp. 459–465, 1994.

[8] Armato S. G., Giger M. L.,.Moran C. J, Blackburn J. T., Doi K., and MacMahon H., "Computerized detection of pulmonary nodules on CT scans," *Radiographics*, vol. 19, pp. 1303–1311, 1999.

[9] Armato S. G., Giger M. L., and MacMahon H., "Automated detection of lung nodules in CT scans: Preliminary results," *Med. Phys.*, vol. 28, pp. 1552–1561, 2001.

[10] Brown M. S., McNitt-Gray M. F., Goldin J. G., Suh R. D., Sayre J. W., and Aberle D. R., "Patient-specific models for lung nodule detection and surveillance in CT images," *IEEE Trans. Med. Imag.*, vol. 20, pp. 1242–1250, Dec. 2001.

[11] Hounsfield GN., "Computed medical imaging", *Med Phys.*, 7(4):283-90, 1980.

[12] Ronse C. and Devijver P. A., Connected components in binary images: the detection problem, Research Studies Press, NY: Wiley, 1984.

[13] Manohar M. and Ramapriyan H. K., "Connected Component Labeling of Binary Images on a Mesh Connected Massively Parallel Processor," *Computer Vision, Graphics, and Image Processing*, vol. 45, 1989, pp. 133-149.

[14] Stefano L. D. and Bulgarelli A., "A simple and efficient connected components labeling algorithm," in *Proceedings of International Conference on Image Analysis and Processing*, 1999, pp. 322-327.

Serhat Özekes was born in Washington D.C. in 1978. He received his B.Sc. and M.Sc. degrees from the Marmara University, Department of Electronics and Computer Education in 2000 and 2002 respectively. He is working for his Ph.D. in the same department. He has been working as lecturer in Istanbul Commerce University since 2002. Recently he is studying on computerized detection of abnormalities in medical images.