



Presented at the Research Congress 2013
De La Salle University Manila
March 7-9, 2013

LUNG NODULE DETECTION AND DIAGNOSIS USING CIRCLE DETECTION THROUGH PLAIN RADIOGRAPHS

Ria Rodette de la Cruz, Trizia Roby-Ann Roque, John Daryl Rosas,
Charles Vincent M. Vera Cruz, Macario Cordel II, and Joel Ilao
Center for Automation Research, De La Salle University-Manila
2401 Taft Avenue, 1004 Manila, Philippines

Abstract: In this paper, we present a system that locates pulmonary nodules in digital chest radiographs through pattern recognition. Digital radiographs that are already diagnosed with lung nodules underwent histogram equalization in order to address varying illumination levels across different regions in the radiographs, and make the radiograph samples more comparable. Laplacian of Gaussian filtering is next applied in order to highlight the edges of pathological features like nodule-shaped blobs in each radiograph. Circular Hough Transform (CHT) was utilized in tandem with pixel-based image processing techniques in locating possible nodules. These system reports the count and sizes of the candidate nodules. We report an overall system accuracy of 73.33% when classifying digitized radiographs as either with nodules or without nodules.

Key Words: Lung Nodules; Pattern Recognition; Feature Extraction; Circular Hough Transform

1. INTRODUCTION

Over the years, medical records in health institutions have recorded lung problems as some of the most common ailments being diagnosed. In the Philippines alone, Acute Lower Respiratory Tract Infection and Pneumonia (Top 1: 690, 566 cases), Bronchitis (Top 2: 616, 041 cases), and Tuberculosis (Top 6: 114, 360 cases) are the leading causes of morbidity from 2000 to 2005, with the average number of patients per 100,000 Filipinos, according to the Philippine Health Statistics of the Department of Health (DoH). These records (Philippine Health Statistics, 2005) are alarming since the number of patients diagnosed with these lung problems continues to rise (Global Health Observatory Data Repository, 2004). Doctors have observed that a lung condition can be associated with a discernible pattern. Thus, it can be said that doctors perform pattern recognition techniques in determining a person's lung condition (Noriyasu, 2009).

Radiologists examine chest x-ray images and this practice is one of the basic and practical procedures in diagnosing lung abnormalities. They apply their medical knowledge in manually recognizing the patterns of lung diseases. The concept of incorporating manual pattern recognition with medical technology is extensively practiced nowadays, as in the case of chest radiography and Computed Tomography (CT-Scan) (Duda and Hart, 1973). Since pattern

HCT-II-015

recognition is done manually, results become subjective and depend on the interpretation of the reader. Moreover, radiologists or medical experts in some rural areas are not readily available to interpret the plain radiographs for the early detection of common lung abnormalities. It is therefore desirable to develop a system that can objectively identify these anomalies in chest plain radiographs. One of the possible solutions is to use image processing for the automatic detection of lung abnormalities. Currently, there are only few works related to automated pattern recognition on digital radiographs (Tonpho, Leelasantithan, and Kiattisin, 1973; Chapelle, Haffner and Vapnik, 1999).

Support Vector Machines (SVM) is used for the classification of lung nodules (Mousa and Khan, 2002). They used several algorithms to enhance the nodule images. After a few tests, a combination of special domain high frequency emphasis filtering (SPD-HFEF), histogram equalization and homomorphic filtering gave the best enhancement among other techniques. The edges were enhanced using SPD-HFEF, the contrast was improved using histogram equalization and the noise was removed using the homomorphic filtering. For feature extraction, they segmented the nodules, performed edge detection and obtained the mean and variances of the resultant edges related to the nodules. Their research achieved an accuracy of 75% using SVM polynomial degree of 10. Another work done (Shimada et al., 1996) was a proposal of a new technique for enhancing pulmonary nodule densities. They developed an eye-shaped filter by adding directionality to the function of the existing difference filter. Their initial detection rate was 79.9% because the filter used could not detect nodules near the thoracic wall. When Hough transformation together with the eye-shaped filter was applied to outline the thoracic wall, the detection rate increased to 87.8% and pulmonary nodules can be detected regardless of their location.

The digital radiographs used in this paper are retrospective and are assumed to have passed the parameters of obliquity, inspiration, and penetration. To ensure fully developed lungs in terms of capacity and avoid inclusion of deteriorating lungs, this study made use of digital chest radiographs of adults, i.e. ages between 18 to 50 years old. For the data testing of the system, digital radiographs were gathered from the De La Salle Health Sciences Institute (DLSHSI). The images were converted from a Digital Imaging and Communications in Medicine (DICOM) format to a Tagged Image File Format (TIFF) which is a lossless image format, considering the memory space that the system will consume in doing its routines. The number of images gathered for Normal and Lung Nodule cases are 53 and 31 respectively.

The system focuses on two lung conditions; Figure 1. a illustrates the Normal case of a chest x-ray image, and Figure 1. b and Figure 1. c illustrates chest x-ray images of the Lung Nodule cases. Regions enclosed in the boxes are the actual locations of the lung pathology.

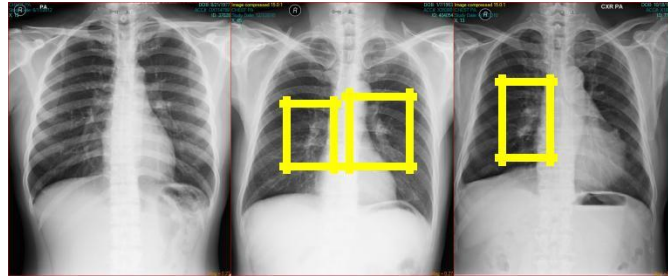


Figure 1. Chest x-ray images of (a) Normal Case, (b) and (c) are both Lung Nodule Cases.

In the Normal case, the lungs are symmetrical and gray, and no nodes or other masses and no collection of air or fluid are visible. The lung should not be smaller or blacker in appearance than the other, and lung markings should not be visible. The diaphragm is a dome-shaped and denser (white) structure compared to the less dense (black) lungs. For the Lung Nodule case, the pattern is seen as an abnormal appearance of solid, elliptical cysts inside the lung area. The number of nodules could exceed one on the left and/or right part of the lungs (Dick, 2000).

High Boost Filter is type of sharpening filter used to enhance high spatial frequency components while suppressing the low spatial frequency components (Chaudhuri, 2009). Equation 1 represents the general equation, wherein high-boosted images are formed by subtracting the low-pass form of the input image from its amplified form. (Bebis, 2001).

$$\begin{aligned}
 \text{Highboost} &= A \text{ Original} - \text{Lowpass} \\
 &= (A - 1) \text{ Original} + \text{Original} - \text{Lowpass} \quad (\text{Eq. 1}) \\
 &= (A - 1) \text{ Original} + \text{Highpass}
 \end{aligned}$$

where:

A = amplification mask

Circular Hough Transform (CHT) is a mapping algorithm that is performed to detect the presence of circular shape or objects in an image. This technique had been used in various researches such as the detection of fingertips position, automatic ball recognition and iris detection for face recognition. CHT is a transformation of a point in the image space to the parameter space defined according to the shape of the object of interest, in this case, a circle. To find circles in an image, the edges are identified and a circle is drawn with the radius r . (Eq. 2-1)

HCT-II-015



shows the equation for a circle (Pedersen, 2007).

$$r^2 = (x - a)^2 + (y - b)^2 \quad (\text{Eq. 2-1})$$

where:

- r = radius of the circle
- x = x-coordinate of a specific point on the circle edge
- y = y-coordinate of a specific point on the circle edge
- a = x-coordinate of the center of the circle
- b = y-coordinate of the center of the circle

(Eq. 2-2) and (Eq. 2-3) shows the parametric representations of a circle. Each edge point in the image adds a circle with the desired radius to an accumulator that has a peak where circles overlap at the centre of the original circle.

$$x = a + r \cos(\Theta) \quad (\text{Eq. 2-2})$$

$$y = b + r \sin(\Theta) \quad (\text{Eq. 2-3})$$

- x = x-coordinate of a specific point on the circle edge
- y = y-coordinate of a specific point on the circle edge
- a = x-coordinate of the center of the circle
- b = y-coordinate of the center of the circle
- r = radius of the circle
- Θ = reference angle in between the x-axis and the radius

2. METHODOLOGY

Figure 2. shows the block diagram for the designed system of classifying lung nodules.

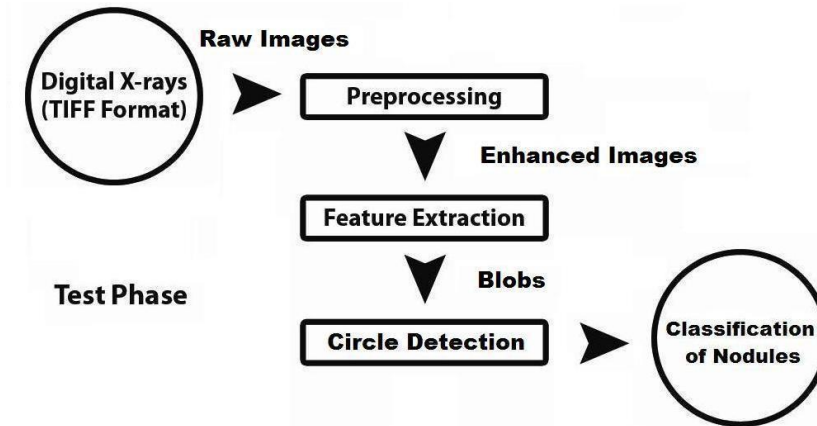


Figure 2. Block Diagram for the Lung Nodule module

The Preprocessing module provides an improved image such that features for extraction in the next block are emphasized. It includes the following techniques in this order: cropping using input coordinates, image resizing, grayscaling, and histogram equalization.

Cropping using input coordinates is the only manual process in the system. The user defines the two points in the image, shown in Figure 3. a, to set the region of interest of the lung image as seen in Figure 3. b. Cropping is followed by *image resizing* which is performed using the Nearest-neighbor interpolation; this sets the image to a square image, standardizing the input images to the system. Conversion to grayscale is then applied to the input images in order to have uniformity in data because x-ray images are usually blue in color. In this study, the histogram of the x-ray image luminance is considered. Thus, *grayscale* process, which normalizes the image into its 8-bit gray-level equivalent, was applied to remove the hue and saturation.



Figure 3. (a) Selection of the Region of Interest, (b) selected region of interest

Once the image is converted to its gray space equivalent, the image undergoes *histogram equalization* to remove the bias to the exposure level of the x-ray images by spreading out intensity values on the histogram. Moreover, it produces better result for region-based feature extraction as this will emphasize the high-dense area (right-skewed histogram) and low-dense area (left-skewed histogram).

The Feature Extraction module acquires the values from regions of interest depending on the technique used. From the system, blobs formed in the lung nodule area are the regions where features would be extracted such as the center of each blob (centroid), and the number of pixels per present blob (area). It includes the High Boost Filter, Disk Filter and Laplacian Filter.

High Boost filter converts the histogram equalized image such that high spatial frequency components are heightened to make the nodules appear in high intensity for easier shape measurement, as shown in Figure 4.



Figure 4. Image Enhancement using (a) Histogram Equalized chest radiograph, and (b) High-Boost Filter.

After getting the high-boosted image, it will be converted into a binary image order to sort the nodule blobs inside the pleural space. Once binarized, the *Disk Filter* is used to shape the blobs by its circular averaging technique to smoothen the edges of the blobs for easier computation of the centroid and the area because of the noise reduction. Lastly, *Laplacian Filter* is applied into the disk filtered image in order for the outline of the nodules to show up due to the high frequency data values from the edges. Figure 5 shows the step-by-step implementation of the two filters on a binary image.



Figure 5. Visual Definition of an (a) Original Image in Binary, (b) Image using Disk Filter, and (c) Image using Laplacian Filter

Circular Hough Transform is a shape analysis method that estimates the parameters of various shapes from its boundary points. With the use of Hough transform, the parameters and location of circles found in an image can be known. This technique makes use of different filters to enhance and outline circular regions within the image, and disregards the coordinates not detected with a circle. Moreover, this process measures the diameters of the found circles.

Plotted output images will be then compared to raw images that have already been diagnosed by radiologists; wherein the location of the nodule in each image have been plotted. To test and verify the performance of the system, manual comparison of the result set from system will be done using the result set from radiologists, following the format of the Receiver Operating Characteristics (ROC) curve.

The ROC curve method is defined by the equations below:

$$\begin{aligned} FPR &= FP \div N \\ &= FP \div (FP + TN) \end{aligned} \quad (\text{Eq. 3-1})$$

where:

FPR = False Positive Rate

FP = False Positive, the number normal chest x-rays classified as abnormal.

TN = True Negative, the number abnormal chest x-rays classified as normal. N = total number of normal x-rays

$$\begin{aligned} TPR &= TP \div P \\ &= TP \div (TP + FN) \end{aligned} \quad (\text{Eq. 3-2})$$

where:

TPR = True Positive Rate

TP = True Positive, the number abnormal chest x-rays classified as abnormal.

FN = False Negative, the number normal chest x-rays classified as normal.

P = total number of abnormal x-rays

$$ACC = (TP + TN) \div (P + N) \quad (\text{Eq. 3-3})$$

where:

ACC = Accuracy

TP = True Positive count

TN = True Negative count

P = total number of abnormal x-rays

N = total number of normal x-rays

$$\begin{aligned} SCP &= TN \div N \\ &= TN \div (FP + TN) \\ &= 1 - FPR \end{aligned} \quad (\text{Eq. 3-4})$$

where:

SPC = Specificity

N = total number of normal x-rays
TN = True Negative count

FP = False Positive count

FPR = False Positive Rate

3. RESULTS AND DISCUSSION

This study used MATLAB for image preprocessing and feature extraction. The data set consists of 60 images in total that were used for the system to process. 30 of which are from the Normal case, and the other 30 are from the Lung Nodule case. The same data set were sent to Dr. Adrian Rabe and Dr. Jun Parungao so they can analyze and plot the locations of the nodule(s) per image. Their diagnoses will be the basis on whether the system has located the nodule in comparison with the plots of the radiologists.

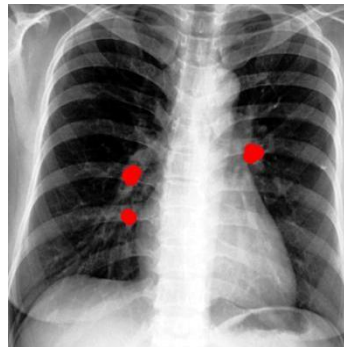


Figure 6. Plotted location of possible lung nodules using CircularHough Transform.

As seen in Figure 6, Circular Hough Transform was used onto the Laplacian-filtered image of the blobs to locate their circular-shaped edges (Young, 2010). TABLE 1. is a representation of the characteristics of the blobs using the image in Figure 6. It includes the centroid location on the x-axis and the y-axis, and the area of the blob by pixel count.

TABLE 1. Nodule Blob Characteristics Data Results for a Nodule image

Blob Number	Centroid Location (x-axis)	Centroid Location (y-axis)	Area (Pixel Count)
1	169.1485	315.5355	478
2	172.6086	388.7391	23
3	182.1153	254.4038	52

To show the classification performance of the system per lung condition, TABLE 2. shows a comparison matrix of the system given the plotted images from the system output and through diagnosis of radiologists. The numbers in the diagonals are the correctly classified lung conditions.

TABLE 2. Comparison Matrix using Diagnosed Raw Images

Lung Conditions		Comparison-based Predicted Class	
		<i>Lung Nodule</i>	<i>Normal</i>
Actual Class	<i>Lung Nodule</i>	25	5
	<i>Normal</i>	11	19

From the table, the difference between the lung nodule and normal cases can be easily identified with 73.33% accuracy. It can be observed that 25 out of 30 images from the lung nodule case and 19 out of 30 images from the normal case were classified correctly, making the system more consistent in categorizing lung nodule-case images. Since the High-Boost filter was utilized, amplifying the areas with high intensity for the nodule-like blobs to be highlighted, some areas in normal-case images with high intensity were also amplified and were also considered as blobs. Thus, more than one-third of the data set for normal case was classified to have lung nodules.

4. CONCLUSION

In this paper, we have presented a comparison-based system that classifies two lung conditions namely: normal, and lung nodule using Circular Hough Transform as feature. For the preprocessing stage, we considered High-Boost filter for nodule enhancements. For the classifier, plotted diagnosed images from radiologists were used for comparison to the images from the system output. The system obtained an overall accuracy rate of 73.33%. Although the accuracy rate of the system is not satisfactory given the standards as cited in some literatures and the lack of a huge data set, the system is open for improvements in the future.

For extensions of this research, an improved system can be achieved by using a larger image database to have better and more reliable training data. Though the accuracy of detecting whether a blob is a nodule or not is still undefined, the group recommends a training and testing phase in order to perform dimension reductions, and determine the characteristic of a blob as a candidate for lung nodule cases.

5. ACKNOWLEDGEMENTS

We would like to thank Dr. Jun Parungao from the Department of Radiology, De La Salle Health Sciences Institute for his help on data gathering, and Dr. Adrian Paul Rabe from



Presented at the Research Congress 2013
De La Salle University Manila
March 7-9, 2013

Department of Internal Medicine, Philippine General Hospital, for his help in the data analysis. Our appreciation is also given to Dr. Lourd Loreto for imparting with the basic medical knowledge about the respiratory system.

6. REFERENCES

- Bebis, G. (2001). *Spatial Filtering*. Retrieved from <https://www.cse.unr.edu/~bebis/CS474/Lectures/SpatialFiltering>.
- Chapelle, O., Haffner, P. and Vapnik, V. "SVMs for histogram-based image classification," IEEE Computational Intelligence Society. Speech and Image Processing Services Research Laboratory, AT&T Labs-Research. Red Bank, New Jersey. September 1999.
- Chaudhuri, P.. (2009). "Sharpening Filters". Department of Computer Science and Engineering, Indian Institute of Technology Delhi. Retrieved December 1, 2012 from Parag Chaudhuri: <http://www.cse.iitd.ac.in/~parag/projects/DIP/assign1/sharp.shtml>
- Dick, E. "Chest x-rays made easy," Student BMJ – The International Medical Journal for Students – Volume 8. September 2000. Retrieved February 12, 2012 from <http://www.e-radiography.net/articles/cxr%20easy%201.pdf>
- Duda, R and Hart, P. "Pattern classification and scene analysis," Medical eBook. Stanford Research Institute. Menlo Park, California. September 1973.
- Global Health Observatory Data Repository. "Cause of Death 2008". World Health Organization. Retrieved June 26, 2012 from http://www.who.int/gho/mortality_burden_disease/global_burden_disease_DTH6_2008.xls
- Global Health Observatory Data Repository. "Global Burden Disease 2004". World Health Organization. Retrieved June 26, 2012 from http://www.who.int/healthinfo/global_burden_disease/DTH6%202004.xls
- Mousa, W. & Khan, M. (2002). *Lung Nodule Classification Utilizing Support Vector Machines*. IEEE Conference Paper - Proceedings of the International Conference on Image Processing. Department of Electrical Engineering, King Fahd University of Petroleum and Minerals. Dhahran, Saudi Arabia.
- Noriyasu, H. "Pattern recognition in medical image diagnosis," Cyberscience center, Tohoku University. Sendai, Japan. October 2009.

HCT-II-015



Presented at the Research Congress 2013
De La Salle University Manila
March 7-9, 2013

Pedersen, S. (2007). *Circular Hough Transform*. Aalborg University, Vision, Graphics, and Interactive Systems . Paper – Proceedings of 2009 International Symposium on Bioelectronics & Bioinformatics. Melbourne, Australia.

Philippine Health Statistics. “Leading causes of morbidity”. Philippine Department of Health. Retrieved February 10, 2012 from <http://www.doh.gov.ph/kp/statistics/morbidity.html#2005>

Seo, N. (n.d.). *Code: Spatial Filters*. University of Maryland, Washington, D.C., U.S.A. Retrieved from Naotoshi Seo: Scientific Softwares or Projects or Experiments on my Graduate Days website: <http://note.sonots.com/SciSoftware/SpatialFilter.html>

Shimada, T., et.al. (1996). *Proposal of a Nodule Density-Enhancing Filter for Plain Chest Radiographs on the Basis of the Thoracic Wall Outline Detected by Hough Transform*. IEICE Transactions on Information and Systems. Nagaoka University of Technology, Nagaoka-shi, Japan.

Tonpho, T., Leelasantithan, A., and Kiattisin, S. “Investigation of chest x-ray images based on medical knowledge and balanced histograms,” International Symposium on Intelligent Signal Processing and Communication Systems. University of Thai Chamber of Commerce. Bangkok, Thailand. December 2010.

Young, D. (2010). *Circular Hough Transform Demonstration*. University of Sussex. Brighton, East Sussex, United Kingdom. Retrieved from MathWorks: http://www.mathworks.com/matlabcentral/fileexchange/26978-hough-transform-for-circles/content/html/circle_houghdemo.html