

Automated Detection of Early Lung Cancer and Tuberculosis Based on X-Ray Image Analysis

KIM LE

School of Information Sciences and Engineering

University of Canberra

University Drive, Bruce, ACT-2601

AUSTRALIA

Abstract: –Early detection of lung cancer and tuberculosis is very important for successful treatment. Diagnosis is mostly based on X-ray images. Our current work focuses on finding nodules, early symptoms of the diseases, appearing in patients' lungs. We use a modified Watershed segmentation approach to isolate a lung of an X-ray image, and then apply a small scanning window to check whether any pixel is part of a disease nodule. Most of the nodules can be detected if process parameters are carefully selected. We are aiming at computerising these selections.

Keywords: –Image segmentation, Watershed segmentation, Medical image analysis, Cancer and tuberculosis, Disease detection, Computer aided diagnosis.

1 Introduction

The treatment of lung cancer and tuberculosis (TB) is easier in the early stages but very difficult in the advanced stages of the diseases. The overall 5-year survival rate for lung cancer patients increases from 14 to 49% if the disease is detected in time [3 cited in 5]. Although Computed Tomograph (CT) can be more efficient than X-ray [5], the latter is more generally available. Therefore preliminary diagnosis for TB and lung cancer, currently performed by medical doctors, is mainly based on lung X-ray images. This is a time-costly process, and the quantity of images to be examined is at an unmanageable level, especially in populous countries with scarce medical professionals.

Computerised analysis of lung X-ray images can reveal these diseases in their early stages. Most cancer and TB cases start with the appearance of small nodules. Our current work aims at the design and implementation of an automated X-ray image analyser to detect early signs of lung cancer and TB.

In image analysis, segmentation is an essential process, which partitions an image into some non-overlapped regions. Image segmentation is often based on the Watershed method [1] or the energy-minimization technique. The Watersnakes method [6], which uses the Watershed as a starting point,

tries to make a link between the two segmentation approaches with the introduction of an adjustable energy function to change the smoothness of the boundary of a segmented area.

This paper reports our continued work in the segmentation of lung areas from X-ray images and the detection of early signs of cancer and TB. The paper is organized as follows. In Section 2, after a brief introduction to the Watershed segmentation, we report some modifications to the approach when it is applied to lung X-ray images. Section 3 presents a method to detect small nodules. The paper ends with a brief discussion.

2 Watershed segmentation

Watershed segmentation was originally used in topography to partition an area into regions. A Watershed segmentation process starts at some regional minima M_i , the lowest points in the area into which water can flow. With an appropriate distance measure, the area is divided into regions Ω_i that are grown from the corresponding minimum M_i by adding to Ω_i , iteratively, unlabelled points on the outer boundary of Ω_i . A point is added to region Ω_i if its distance from the region is smaller than those from other regions. The addition is repeated until no

more point can be assigned to any region. The remaining unlabelled points are those of the watershed line.

In implementation, to obtain a thin-watershed line, a point is added to the region Ω_i even when its distance from the region equals those from some other regions. Hence, there is no point belonging to the watershed line. A modification to the Watershed segmentation, the Watersnake approach, is used to adjust the smoothness of regions' boundaries. In [8], we introduced a new distance measure, the gravity measure, and used it in the Watersnake method to segment a lung X-ray image. The result is illustrated in Figs. 1&2. Due to interference by the ribs, the grey levels of the lung on the left boundary are higher than those on the right. As a consequence, the lung part on the left is constricted, and adjustment is needed to obtain a boundary nearer to the true one.



Fig. 1: Original lung X-ray image



Fig. 2: Segmentation of lung X-ray image using gravity Watersnake method

Watershed segmentation with adjustment

In the current work, we apply the Watershed method with the following steps:

- a) Assume that the lung in the examined X-ray image is surrounded by brighter background. Make a picture frame that totally encloses the right (or left) lung; this condition will be relaxed later with some penalty.
- b) Find the grey-level histogram of the image part within the picture frame. Record four different grey levels: $GL(m)$, $m = 0$ to 3, for 20, 40, 50 and 70%, respectively, of the total number of pixels, e.g., 20% of pixels within the picture frame have grey levels smaller than $GL(0)$.
- c) Divide the picture frame into 4x4 blocks with block (0, 0) being the top-left corner. Sow seeds for the lung object in blocks (1,1), (1,2), (2,1)

and (2,2) at pixels darker than $GL(0)$, and seeds for the background in blocks (1,0) , (2,0), (1,3) and (2,3) at pixels having grey levels greater than $GL(3)$.

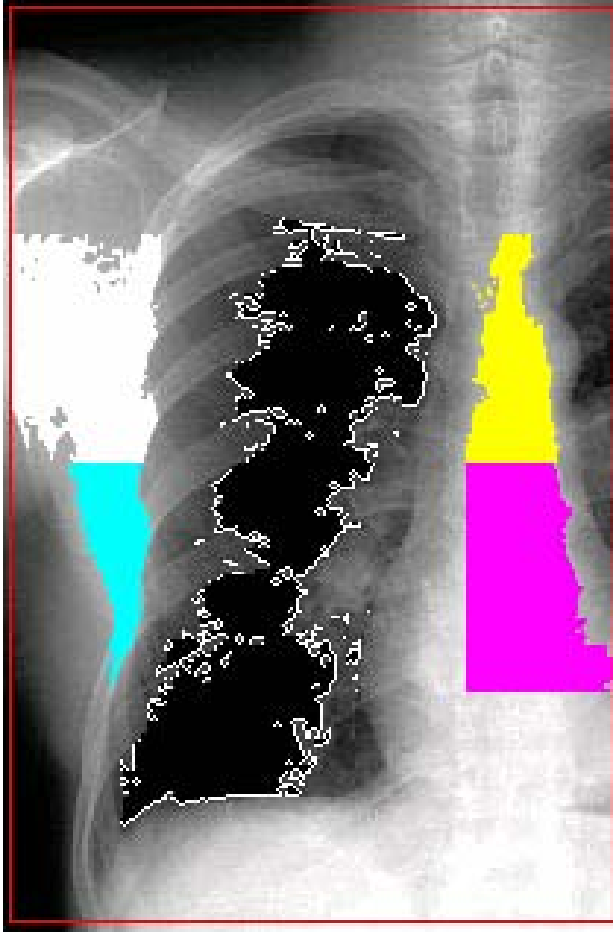


Fig. 3: Core parts of lung and left/right background before Watershed segmentation

- d) Iteratively expand the core of the lung object by including neighbouring pixels that have a brightness less than $GL(0)$, and expand the cores of the background with neighbours that have grey levels greater than $GL(3)$. We obtain a result as shown in Fig. 3. The four background cores play the roles as minima (in the complement image) of four regions $\Omega_{(i, j)}$. This step assumes that the lung and the background areas are respectively greater than 20% and 30% of the picture frame.
- e) Apply the Watershed approach on the four background regions until the region $\Omega_{(1,0)}$ touches

the region $\Omega_{(1,3)}$ on the top of the lung, and the two regions $\Omega_{(2,0)}$ and $\Omega_{(2,3)}$ meet each other at the bottom-left corner of the lung. The touching points are guaranteed by the condition in Step a). Record the grey level when the Watershed process is complete. Use this level as a threshold to expand the lung. We obtain a result as shown in Fig. 4. The lung boundary (Fig. 5) is similar to that previously obtained (Fig. 2).



Fig. 4: Segmented lung X-ray image with a modified Watershed segmentation

- f) Calculate the differential grey level (called SlopeX) in the horizontal direction of every pixel using a Sobel operator [2].
- g) Adjust the left boundary of the lung by slowly increasing the grey-level threshold found in Step e). In each increment, add to the lung object pixels that have negative SlopeX. The iteration process is stopped when in an iterative loop there is not any added pixel or when the threshold is greater than $GL(3)$. This adjustment is based on

the assumption that the left background has rather bright pixels near the lung. The boundary is then smoothed with several (e.g. 5) dilation operations followed by the same number of erosions. The result is presented in Fig. 6.

h)

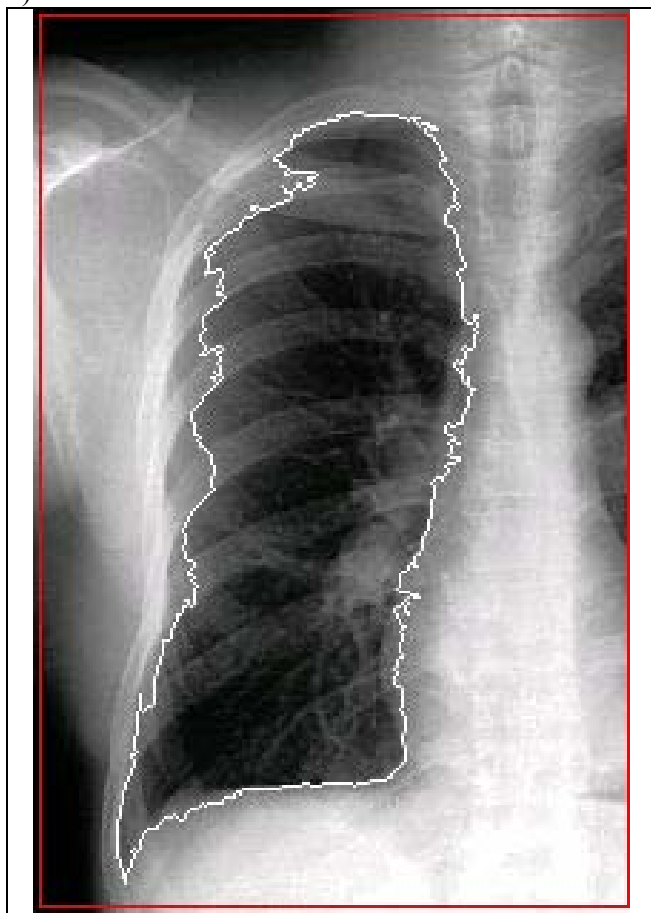


Fig. 5: Lung boundary after Watershed segmentation

Relaxation on the enclosure of lung with picture frame

In Step a) of the above Watershed segmentation application, we assume that the picture frame is big enough so that the lung is totally surrounded by brighter background. This condition is necessary for the selection of the grey-level threshold in Step e). If this condition is not satisfied, the Watershed process will not stop until the background parts are expanded to the grey level $GL(1)$. As a consequence, the right boundary is constricted further inside of its true one as shown in Fig. 7, and adjustment may be needed.

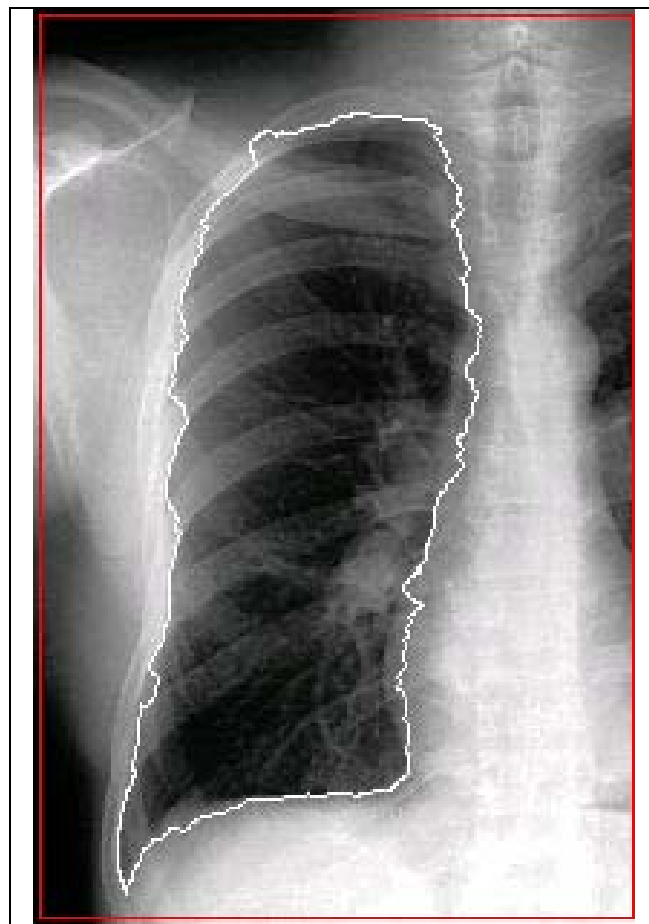


Fig 6: Lung with adjusted left boundary

3 Nodule Detection

The appearance of small nodules is one of the early signs of lung TB and cancer. Nodule pixels are often brighter than the surrounding areas but in some cases, the difference in grey levels is not significant. Furthermore, ribs and pulmonary arteries, which often have higher grey levels, also contribute to the complexity of lung tissue and make some nodules undetectable. In up to 30% of cases, nodules are overlooked by radiologists on their first examinations [7], although they are visible in retrospect, especially when computer-aid diagnostic tools are used to focus radiologists' attention on suspected areas [4].



Fig. 7: Right lung boundary is constricted inside of the true one when the background within the picture frame does not totally enclose the lung.

To detect early nodules we use the following approach:

- a) Apply a small fixed size window –called scanning window– to every pixel (inside the picture frame), which has not been marked as part of suspected nodules.
- b) Find the average and the maximal grey levels of the pixels within the scanning window. Select a local grey-level threshold between the average and the maximal levels.
- c) Count the number of pixels that have grey levels higher than the local threshold. If the counted number is within a specific range then mark the pixel as part of a suspected nodule.

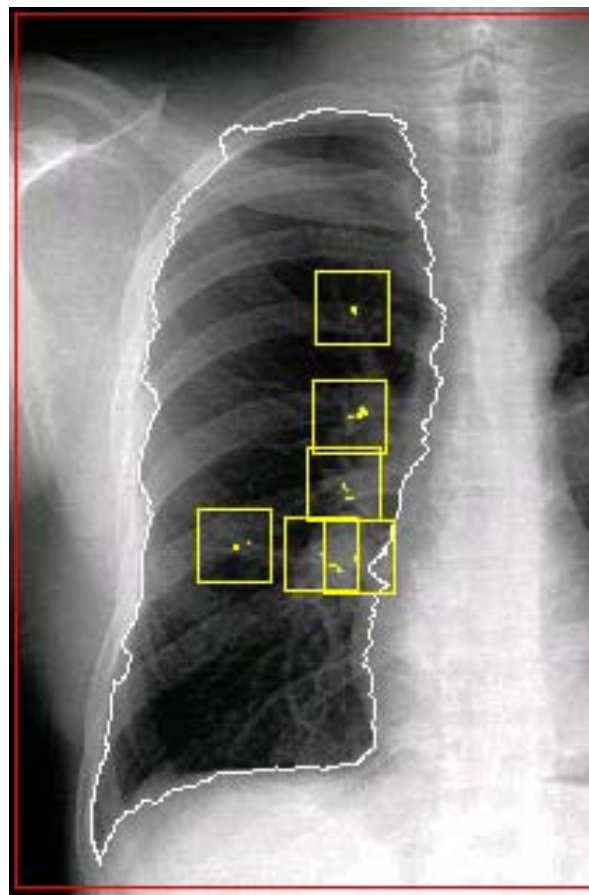


Fig. 8: Detection of nodules by applying a small scanning window to all pixels of the lung

Our preliminary selections are as follows. The window size is (41×41) , the grey-level threshold equals $(0.7 \text{ Maximal level} + 0.3 \text{ Average level})$, the number of suspected pixels is in the range between 0.5 to 10% of the total number of pixels in the scanning window area. With these selections, most suspected nodules in the example image and some others are found as shown in Fig. 8. Larger scanning windows may be used for the cases of advanced nodules. With a lower value of the grey-level threshold, fading nodules can be detected as well as false-positive errors. The range of nodule sizes also affects the quality of nodule detection. The lower end of the range is set to reduce false positive errors due to image noise, and the higher end is to exclude ribs and pulmonary arteries. There are correlations among these selections. How to choose these values automatically is a problem that needs further investigation. Some soft computing techniques such

as Fuzzy logic, artificial neural network, Genetic algorithm, etc. are prospective tools for this problem [5].

4 Discussion

In this paper, we present some modifications to the Watershed method to segment lung X-ray images. The boundary of a segmented lung is closer to its true one than that presented in our previous report. We also introduce an approach to detect small nodules, symptoms of early cancer and TB. The experimental results are preliminary so further investigation is needed for the automated selection of process parameters in general cases.

Acknowledgement

This work was performed during an Outside Study Program at the School of Electrical and Information Engineering, University of Sydney, from April to June, 2006. The author is grateful to the Division of Business, Law, and Information Sciences of the University of Canberra for its financial support, the School of Electrical and Information Engineering, University of Sydney, and its staff for their permission to use facilities during the program, especially Dr. Peter Nickolls and Dr. Hansen Yee.

Medical advice from Dr. Peter Nickolls, Dr. Quoc-Truc Nguyen (Ulcer and Cancer Hospital, Saigon) and Dr. Ngoc-Thach Tran (Tuberculosis Hospital, Saigon) is also acknowledged.

Reference:

- [1] Beucher, S. and Meyer, F., "The Morphological Approach of Segmentation: The Watershed Transformation," *Mathematical Morphology in Image Processing*, E. Dougherty, ed., pp. 43-481, New York: Marcel Dekker, 1992.
- [2] Gonzalez R.C. and Woods R.E., "Digital Image Processing", Addison-Wesley Publishing Co, 1992, Chapter 7 (Image Segmentation), pp.413-478.
- [3] Gurcan, M.N., et al., "Lung nodule detection on thoracic computed tomography images: preliminary evaluation of a computer-aided diagnosis system", *Med Phys* 2002, 29(11): 2552-8.
- [4] Kakeda, S. et al., "Improved Detection of Lung Nodules on Chest Radiographs Using a Commercial Computer-Aided Diagnosis System", *American Journal of Roentgenology*, 182, February 2004, pp. 505-510.
- [5] Lin, D.T. et al., "Autonomous detection of pulmonary nodules on CT images with a neural network-based fuzzy system", *Computerized Medical Imaging and Graphics*, 29 (2005), pp.447-458.
- [6] Nguyen, H. T., et al "Watersnakes: Energy-Driven Watershed Segmentation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 25, Number 3, pp.330-342, March 2003.
- [7] Suzuki K., et al., "False-positive Reduction in Computer-aided Diagnostic Scheme for Detecting Nodules in Chest Radiographs by Means of Massive Training Artificial Neural Network", *Academic Radiology*, 12, No 2, February 2005, pp. 191-201.
- [8] Watman, C. and Le, K., "Gravity Segmentation of Human Lungs from X-ray Images for Sickness Classification", *Fourth International Conference on Intelligent Technologies (InTech'03)*, 2003, pp. 428-434.