Automated Detection of Lung Tuberculosis Based on X-ray Image Analysis

May Thitsa Oo, Thin Thin Htway
Computer University (Hinthada)
maythitsaoo@gmail.com

Abstract

Detection of lung tuberculosis is mostly based on X-ray images. Image segmentation is important in different fields of image processing. Image segmentation is process of dividing images according to its characteristics. Different methods are presented for image segmentation. In this paper to find nodules, symptoms of diseases in X-ray images, we use watershed segmentation approach. First take the gray scale image and then applying the watershed segmentation approach to segment the image with catchment basins. When the lung image is isolated from X-ray image, the suspected nodule pixels in the lung can be found. Based on nodules and user's input symptoms, patient can be defined either suffer from lung TB or not.

Keywords: watershed transformation, catchment basins, lung tuberculosis.

1. Introduction

Tuberculosis (TB) describes an infectious disease that has plagued humans since the Neolithic times. Tuberculosis (TB) is a common and often deadly infectious disease caused by a germ, called *Mycobacterium tuberculosis.* TB usually attacks the lungs but can also affect other parts of the body. TB symptoms may not appear until the disease has already caused damage. The classic symptoms are chronic cough, fever, loss of appetite, night sweats and weight loss. Tuberculosis is difficult to diagnose because the signs and symptoms are similar to those caused by other diseases. Treatment is difficult and requires long courses of multiple antibiotics. Preliminary diagnosis of lung TB is mainly based on lung X-ray images. Most TB cases start with the appearance of small nodules on lung's surface. Our current work aims at the design and implementation of an automated X-ray image analyzer to detect signs of lung TB.

Image segmentation is a process that partitions an image into its constituent regions or objects according to its characteristic e.g., color and objects present in the images. These regions are sets of pixels. The result of image segmentation is in the form of images that are more meaningful, easier to understand and easier to analyze. Segmentation should stop when the objects of interests in an application have been isolated. For detection of nodules in lung's X-ray image, there are many image processing methods for image segmentation. This paper focuses on watershed segmentation approach. Watershed segmentation is a morphological based method for image segmentation.

2. Related Area and Problem Issues

M. Couprie introduced "one-dimensional" topology for grayscale image [2]. They precisely define a topological grayscale transformation that generalizes the action of a watershed transformation. Then, proposed an efficient algorithm to compute the topological grayscale transformation. L. Najman proposed watershed of a continuous function [6]. Using continuous definitions, it presents the watershed differences with classical edge detectors. In which, watershed algorithms present the advantage of allowing the use of makers and\or anchor points, thus opening the ways towards grey-tone skeletons. J.B.T.M. Roerdink and A. Meijster presents the details of watershed transform, definition and parallelization strategies [7]. It explains differences between watershed transforms based on different definitions with various examples.

Morphological watershed algorithms are used in detection of In Vivo Early cancer and Breast cancer tumor [3, 5]. To avoid over-segments, In Vivo Early Cancer Detection use the more advanced marker-controlled watershed transform. M.S.H. Khiyal gives some basic information about watershed segmentation algorithm on 2D images [4]. It uses MATLAB tools. S. Beucher introduces image segmentation by mathematical morphology based upon the notions of watershed through various examples of segmentation in materials sciences, electron microscopy and scene analysis [1].

3. Image Processing

Image processing is a physical process used to convert an image signal into a physical image. The image signal can be either digital or analog. The actual output itself can be an actual physical image or the characteristics of an image.

A digital image is a representation of a two-dimensional image as a finite set of digital values, called picture elements or pixels. Pixel values typically represent gray levels, colors, heights, opacities etc. Digitization implies that a digital image is an approximation of a real scene. Digital image processing focuses on two major tasks:

(1) Improvement of pictorial information for human interpretation.
(2) Processing of image data for storage, transmission and representation for autonomous machine perception.

3.1 Image segmentation

Segmentation is often the most difficult problem to solve in image processing. Segmentation refers to the process of partitioning a digital image into multiple segments (sets of pixels, also known as superpixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics.

Segmentation algorithms for monochrome images generally are based on one of two basic properties of gray-scale values:

- Discontinuity: The approach is to partition an image based on abrupt changes in gray-scale levels. The principal areas of interest within this category are detection of isolated points, lines, and edges in an image.
- Similarity: The principal approaches in this category are based on thresholding, region growing, and region splitting/merging.

3.2 Watershed transformation

The watershed transformation is an efficient tool for segmenting grayscale images. The watershed constitutes one of the main concepts of Mathematical Morphology. The watershed transform can be classified as a region-based segmentation approach. Watershed segmentation gets its name from the manner in which the algorithm segments regions of an image or volume into catchment basins.

Any grayscale image can be considered as topographic surface (Figure.1). Topographical interpretation consists of three points. They are points belonging to regional minimum, catchment basins and divide lines or watershed lines [8]. Imagine the surface of image being immersed in water; catchment basins will be filled up with water starting at each local minimum. A minimum can be considered as a sink of the topographic surface. Dams are built to prevent the rising water from distinct catchment basins from merging. Only the tops of the dams are visible above the water lines. These dam boundaries correspond to the divide lines of the watersheds.

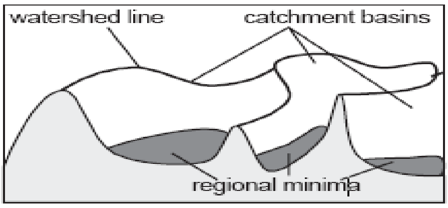


Figure 1: Topographic surface

3.2.1 Watershed Definition: A watershed definition for the continuous case can be based on distance function. Assume that the image *f* is an element of the space C (D) of real twice continuously differentiable functions on a connected domain D with only isolated critical points (pixels). Then the topographical distance function between two points (p and q) is according to [7] defined as:

$$T_f(p,q) = \inf_{\substack{\gamma \\ \gamma\in[p\to q]}} \int_\gamma \left\| \nabla f(\gamma(s)) \right\| ds \, ,$$

Where $[p \to q]$ denotes all possible paths from p to q. The path with shortest T_f-distance between p and q is a path of steepest slope.

Topological watershed involves computing for each pixel the minimum pixel at which a drop of water would end up if it fell on the pixel under consideration. Once these values have been computed for every pixel in the gradient image, the algorithm groups pixels with common minima into catchment basins. The catchment basin CB (m_i) is the set of pixels such that water drop falling at p flows down along the relief, following a certain descending path, and eventually reaches m_i.

$$CB(m_i) = \{x \in D \mid \forall j \in I \setminus \{i\} : f(m_i) + T_f(x, m_i) < f(m_j) + T_f(x, m_j)\}$$

The watershed of *f* is the set of points which do not belong to any catchment basin:

$$Wshed(f) = D \cap \left(\bigcup_{i\in I} CB(m_i) \right)^c$$

4. System Design

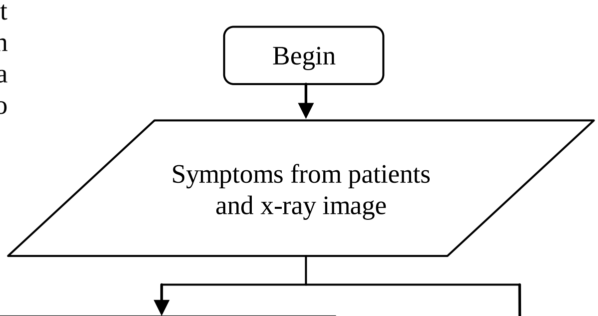
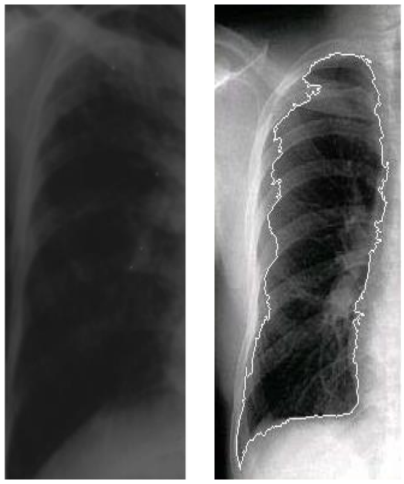


Figure 2: System design for detection of lung TB

5. Implementation Steps

First, scanning the x-ray image of patient with grey scale mode and input the gray scale image to the system. The gray scale image is calculated the catchment basins and combine the catchment basins by using watershed approach. When the watershed process complete, set a grey-level threshold. Iteratively expand the core of the lung object by including neighboring pixels that have brightness less than grey-level threshold. The iteration process is stopped when in an iterative loop there is not any added pixel. When the lung object is enclosed by boundaries, small nodules in the lung can be detected. The result of image segmentation is defined as pre-result for diseases. The inputted symptoms are matching with threshold data from database. Based on the symptom result and x-ray image result, the final result is come out to the screen whether the patient is suffered from lung TB or not.



(a) (b)

Figure 3: (a) Gray scale lung image. (b) Lung boundary after watershed segmentation.

6. Nodule detection

The appearance of small nodules is one of the signs of lung TB. Nodule pixels are often brighter than the surrounding areas and they have higher gray levels. Apply a small fixed size window with size (41×41) - called scanning window- to every pixel. Find the average and maximal grey levels of the pixels within the scanning window. Select the local grey-level threshold between the average and maximal levels. Count the number of pixels that have grey levels higher than the local threshold. If the counted number is in the range between 0.5 to 10% of the number of pixels in the scanning window area, these pixels can be defined as part of suspected nodules (Figure.4). The nodule result is defined as pre-result for diseases. Patient enters symptoms as input. Depend on the x-ray input and symptom input, get the final result.

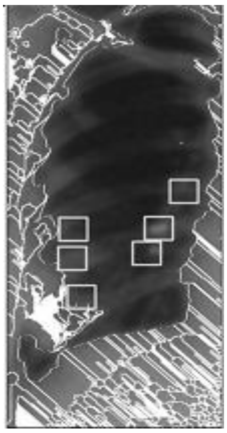


Figure 4: Detection of nodules

7. Simulation Results

In patient data entry form, Patient fills information and selects symptoms they suffered. This information is saved (Figure.5). Then select patient's x-ray image. After the x-ray image is segmented using watershed transformation method, the lung object is enclosed by boundaries. Then nodules are detected. Depending on the number of suspected nodules and patient's suffered symptoms, gets the resultant percentage of TB suffering. Figure (6) shows detection of nodules and final result for lung TB.

Figure 5: Patient data entry

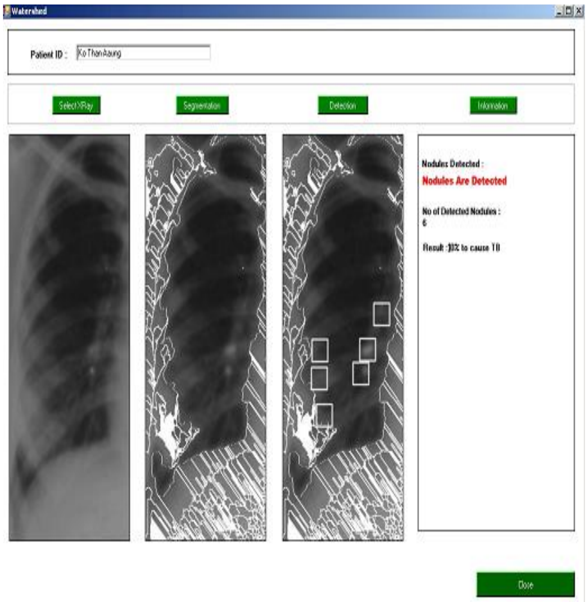


Figure 6: Result for detection of lung TB

8. Conclusion

Watershed segmentation is an effective method for gray level image segmentation. This system presents some modifications to the watershed method to segment lung x-ray images. The most important step is image acquisition. Any deficiency during the image acquisition can cause many problems in the result. The x-ray image needs to be scanned with gray scale mode and store in image library. Applying watershed segmentation approach, the boundary of a segmented lung is closer to its true one. This system is expected to be useful software for patients and experts for helping diagnosing diseases.

9. References

[1] S.Beucher, "The watershed Transformation Applied to Image Segmentation", Centre de Morphologie Mathematique, Ecole des Mines de Paris, 35, rue Saint-Honore 77305 FORTAINEBLEAU CEDEX (France).

[2] M.Couprie and G.Bertrand, "Topological Grayscale Watershed Transformation", SPIE Vision Geometry VI Proceedings, Vol.3168, pp.136-146, 1997.

[3] M.K.D.Ho, 'Watershed Segmentaion Algorithm for Medical Confocal Image Analyses towards In Vivo Early Cancer Detection", Department of Biology, Johns Hopkins University.

[4] M.S.H.Khiyal, A.Khan, and A.Bibi, "Modified Watershed Algorithm for Segmentaion of 2D Images', Computer Sciences Dept., Fatima jinnah women University, Rawalpindi, Pakistan.

[5] A.Kandaswamy and H.S.Sheshadri, "Detection of Breast Cancer Tumor based on Morphological Watershed Algorithm", Department of ECE, PSG College of Techonology, Coimbatore-641004.

[6] L.Najman and M.Schmitt, "Watershed of a continuous function", Signal Processing 38(1994) 99-112.

[7] J.B.T.M.Roerdink and A.Meijster, "The Watershed Transform: Definition, Algorithms and Parallelization Strategies", Institude of Mathematics and Computing science, University of Groningen.

[8] A.Verma (Y6120), V.Annapurve (Y6088), "Watershed Image Segmentation".