Image processing based detection of lung cancer on CT scan images

Bariqi Abdillah, Alhadi Bustamam, and Devvi Sarwinda
Faculty of Mathematics and Science, Universitas Indonesia, Depok, Indonesia
E-mail: \{bariqi.abdillah, alhadi, devvi\}@sci.ui.ac.id

Abstract. In this paper, we implement and analyze the image processing method for detection of lung cancer. Image processing techniques are widely used in several medical problems for picture enhancement in the detection phase to support the early medical treatment. In this research we proposed a detection method of lung cancer based on image segmentation. Image segmentation is one of intermediate level in image processing. Marker control watershed and region growing approach are used to segment of CT scan image. Detection phases are followed by image enhancement using Gabor filter, image segmentation, and features extraction. From the experimental results, we found the effectiveness of our approach. The results show that the best approach for main features detection is watershed with masking method which has high accuracy and robust.

1. Introduction
Cancer is one of most dangerous disease that causes deaths [1]. Data obtained from Global Burden Cancer shows that in 2012 there are 14.1 million cases of cancer in the world, with lung cancer occupies the first position with a percentage of 13%. While the number of cancer deaths was recorded 8.2 million deaths, with lung cancer cause of death in first place with a percentage of 19% [2]. Lung cancer is a disease of abnormal cells multiplying and growing into a tumour. Cancer cells can spread out from the lungs through the blood stream or lymph fluid that surrounds the lung tissue. Generally, cancer cells often spread toward the center of the chest due to the natural flow of lymph. Metastasis occurs when cancer cells spread to other organs. The process of early detection of cancer plays an important role to prevent cancer cells from multiplying and spreading. Previous researches have been conducted for analyzing lung cancer such as using clustering method in microarray data [3], the detection of lung cancer with general image processing techniques in CT scan data with good results and accuracy [4]. In this study we proposed and evaluated additional image segmentation methods in analyzing lung cancer using image processing techniques.

Figure 1 shows a general description of lung cancer detection system that contains four basic stages. There are four stages to determine whether there is a lung cancer or not. The first phase is we get CT scan image data. The second phase, we implement image enhancement to improve quality of image. The third phase is image segmentation which is an important step in the detection of cancer. The fourth stage is feature extraction that is give us a conclusion whether there is a lung cancer or not.

2. Methodology
2.1 Image Acquisitions
The imagery used is from CT scan database. Our data is from the VIA and ELCAP Public Access
Figure 1. Image Processing for Lung Cancer Detection Stages

Database [5]. Image is in digital imaging and communications in medicine (dicom) format. There are 50 patients and each patient have about 250 pieces of images. Figure 2 shows example of CT scan image data.

Figure 2. Example three pieces of a patient

2.2 Gabor Filter

Gabor filter named by Dennis Gabor, is a linear filter used for edge detection. Representation of Gabor filter similar to the human visual system. In the spatial domain, 2D Gabor filter is a Gaussian filter function modulated by a sinusoidal function [4]. In the process of this cancer detection imagery used is a 2D image, so using 2D Gabor filter.

\[
G(\sigma, \theta, \lambda, \psi; x, y) = \exp\left(-\frac{(x^2 + \gamma^2 y^2)}{2\sigma^2}\right) \cdot \cos\left(\frac{\lambda}{\sigma} \cdot x + \psi\right)
\]

Figure 3 represents the input image and the results after enhanced by Gabor Filter.

2.3 Region Growing

Region growing is a procedure to extend the covered area by collecting pixels or sub-regions based on predetermined criteria. Basically, the approach is to initiating a set point (seed) and then expanding the area surrounding the seed which has the same properties as the seed, for example, the area that has the same range of color or gray level [6]. Figure 4 shows three stages of segmentation by region growing method.

2.4 Marker Controlled Watershed

There are two main approaches in segmentation, including edge approach and regional approach. Watershed segmentation method combining both of this approaches. Watershed method is a powerful
method to get fast segmentation results [7]. The basic idea of segmentation watershed transformation comes from geography, where an image is viewed as an earth surface (topography) with gray level as a measure surface height. Rainwater will flow and fill the lowest surface of catchment basins [6]. Figure 5 is the stages of watershed segmentation.

![Image](image1.png)

**Figure 3.** Image input and output by Gabor Filter

![Image](image2.png)

**Figure 4.** stages of region growing method

![Image](image3.png)

**Figure 5.** stages of watershed segmentation method
3. Results

3.1 Region Growing
The process to get the segmentation by region growing method is as the following; firstly, select the area that will be the target object, which are the right lung and the left lung, then, put the seed in this area. Furthermore enlarge the size of the seed so that it covers all the desired areas. Figure 6 shows the result of segmentation by region growing method.

![Figure 6. segmented by region growing method](image)

3.2 Marker Controlled Watershed
To get the segmentation result of watershed method, the steps are as follows; firstly, calculate the distance gradient for edge detection, and then mark the target object by using a morphological technique called opening by reconstruction and closing by reconstruction. After marking object is revealed, other areas can be discarded. Modules and example of marker controlled watershed segmentation method is existing in Matlab software that can be directly used. Figure 7 shows the result of segmentation by using Marker-Controlled Watershed.

![Figure 7. segmented by marker controlled watershed method](image)

3.3 Marker Controlled Watershed with Masking
Segmentation by this method is the same process as the segmentation method at subsection 3.2 with additional masking in this process, which marks the area that contains the target object at the right and the left lung. Results of segmentation using Marker-Controlled Watershed with Masking is shown in Figure 8.

3.4 Three Methods Segmentation Analysis
The analysis of three segmentation method based on their success to find the segmentation results of the target object. Figure 6 shown two method have successfully found the target object including (a) region growing and (c) marker controlled watershed with masking. meanwhile (b) marker controlled watershed segmentation has fail to get segmentation results, since there are background surrounding the target object.
Figure 8. segmented by marker controlled watershed with masking method

Figure 9. comparing three output segmentation method

Table 1 shows the comparison of running time of region growing method and marker controlled watershed segmentation with masking method. Marker controlled watershed method (b) is not included to the comparison because this method already fail to get segmentation results.

| Image | Region Growing | Marker Controlled Watershed with Masking |
|-------|----------------|------------------------------------------|
| 1     | 93.259 s       | 4.644 s                                  |
| 2     | 88.136 s       | 4.639 s                                  |
| 3     | 87.740 s       | 4.329 s                                  |
| 4     | 87.870 s       | 4.719 s                                  |
| 5     | 89.316 s       | 4.470 s                                  |

Next analysis is based on the processing time required to obtain segmentation results. Region growing takes 90 seconds on average to get the segmentation result, while the marker controlled watershed by masking is takes 5 seconds only to produce the image segmentation results. So we choose to use marker controlled watershed in this study.

3.5 Binarization

Binarization is the process of changing the color of the pixel values into two classes [8], such as black and white. After getting the quantity of black and white pixels on segmentation results, then we compared it with a threshold value to determine the condition of lung (normal or cancer). The threshold value is obtained from observations on normal lung. The threshold value that is used in this research is 17178.48. If the number of black pixels more than the threshold, then we conclude the lung is normal, otherwise
the lung is cancer.

![Diagram showing pixelization and binarization scheme]

**Figure 10.** Pixelization and binarization scheme

Table II shows result of detection from five sample patients.

**Table 2. Detection result of five sample patients.**

| Patient | Black pixels | Conclusion |
|---------|--------------|------------|
| 1       | 28504        | normal     |
| 2       | 26661        | normal     |
| 3       | 18673        | normal     |
| 4       | 10399        | cancer     |
| 5       | 13865        | cancer     |

4. Conclusion

In this study we implement and evaluate three image segmentation methods for analyzing lung cancer, such as Region Growing, Marker Controlled Watershed, and Marker Controlled Watershed with Masking. The results show that Marker Controlled Watershed with Masking give us the best performance in term of segmentation result and running time. Therefore, we select Marker Controlled Watershed with Masking method in image segmentation stage. Furthermore, in the feature extraction stage, we use color attribute for the analysis of lung cancer using binarization. Finally, the binarization method was successfully determined condition of lung (normal or cancer) from the CT scan image.

Acknowledgment

This research is supported by HIBAH PITTA UI 2016 research grant.

References

[1] Permata T S and Bustamam A 2015 Clustering protein-protein interaction network of TP53 tumor suppressor protein using Markov clustering algorithm *International Conference on Advanced Computer Science and Information Systems (ICACSIS)* pp 221–226

[2] GLOBOCAN 2012, International Agency for Research on Cancer, World Health Organization. http://globocan.iarc.fr/

[3] Khalid R 2014 Clustering analysis of cancerous microarray data *Journal of Chemical and Pharmaceutical Research* 6(9) pp 488-493, ISSN: 0975-7384

[4] Kajal N et al 2015 Early Detection of Lung Cancer Using Image Processing Technique: Review *International Journal of Advent Research in Computer and Electronics (IJARCE)* 2(2), E-ISSN: 2348-5523
[5] Lung Cancer Database, Available at: https://eddie.via.cornell.edu/cgibin/datac/signon.cgi  
[6] Gonzalez R C and Woods R E 2008 *Digital Image Processing Upper Saddle River* (New Jersey: Prentice Hall)  
[7] Allaoui A E and Nasri M 2012 *Medical Image Segmentation by Marker Controlled Watershed and Mathematical Morphology* 1LABO MATSI, ESTO, B.P 473, University Mohammed I OIJDA, Maroko  
[8] Sarwinda D and Bustamam A 2016 Detection of Alzheimers disease using advanced local binary pattern from hippocampus and whole brain of MR images *International Joint Conference on Neural Networks (IJCNN)* IEEE pp 5051–5056