# Automatic Lung Cancer Detection Using Color Histogram Calculation

Ratih Wulandari
Department of Informatics and Computer
Electronic Engineering Polytechnic Institute of Surabaya
Surabaya, Indonesia
ratihwul@ce.student.pens.ac.id

Riyanto Sigit, Setia Wardhana
Department of Informatics and Computer
Electronic Engineering Polytechnic Institute of Surabaya
Surabaya, Indonesia
riyanto@pens.ac.id, setia@pens.ac.id

***Abstract*—Lung cancer is a disease that caused by uncontrolled cell growth in lung. Lung cancer is still the first worldwide killer. CT Scan Thorax is a method for early detection of lung cancer patients. However, cancer detection in lung CT-Scan image still done manually. In this paper, the segmentation of lung image is proposed. Cancer segmentation will process the lung CT-Scan as an image input with watershed process to cut off cavity area. The result will be processed by color histogram calculation to obtain mean and standard deviation value. This value is useful for evaluate non-cancer area and produce cancer image. Segmentation process will be followed by measurement of cancer and cavity area. The overall output is percentage between the large of cancer area and cavity area. The experiment represented that this method is able to detect lung cancer automatically. The performance segmentation for assessment errors obtained an average cavity area segmentation 12.75% and cancer area segmentation 31.74%.**

***Keywords—CT-Scan; lung; watershed; histogram; mean; standard deviation***

## I. INTRODUCTION

The first diagnosis for lung cancer commonly use CT-Scan [5]. Lung CT-Scan is a method of medical imaging using tomography to produce a three-dimensional image of the inside of the lungs with taking in one round (axis) [5]. CT-Scan works by reflecting x-ray beam [8]. CT-Scan images are formed by gray values in different ranges [4]. Hard tissue (such as bone) will be lighter than soft tissue (such as muscle). Area without tissue (air) will have black color.

Research will be continued on automatic lung cancer detection. Lung CT-Scan image input will be processed to obtain cancer image with percentage calculation of cancer. The author uses watershed transformation to eliminate area in order to get inner objects in the lung cavity. Watershed transformation algorithm is usually implemented by simulating a flooding process (Hagyard et al., 1996; Vincent and Soille, 1991). Process will continued by using color histogram due to different gray value factors between cancer object and other objects. Histogram collects the values from the selected channels in the input images and finds the corresponding histogram bin. Output is data with a predetermined gray value, which is the gray value of object cancer.

## II. METHODOLOGY

The procedure for cancer segmentation which involves erosion and dilation for preprocessing. Image will be processed by watershed to crop cavity area. Image result will be calculated by histogram calculation, mean and standard deviation to segment cancer area. The procedure for cavity segmentation is drawn from the watershed process through binary process then noise reduction using median filter. The procedure are graphically described in a schematic diagram shown in Fig. 1.

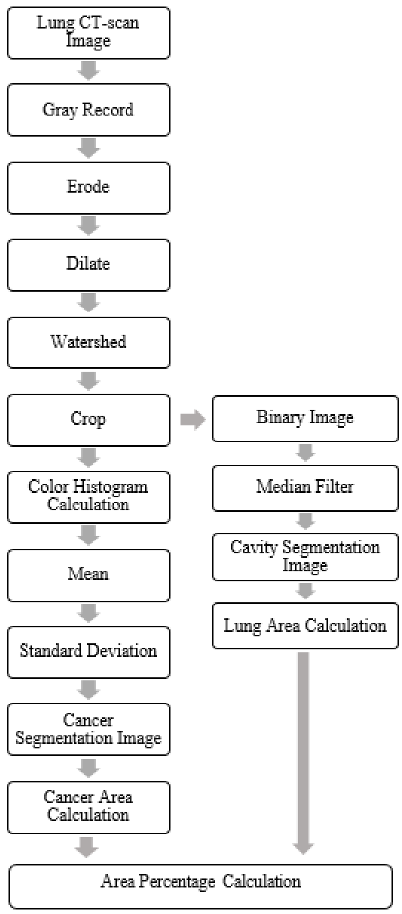



Fig. 1. Schematic diagram of the algorithm

*A. Lung CT-Scan Image*

Almost all CT-Scans were performed in axial plane [8]. The axial plane is selected for clearer image results. CT-Scan produces grayscale images with different gray levels as shown in Fig.2 [4].

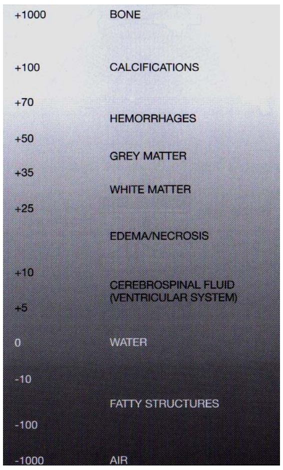



Fig. 2. Gray level of CT-Scan image

In this research, we used axial plane lung CT-Scan image. Size of input image is 512 x 512 pixels. Fig. 3 represents the lung images in axial plane view.

Fig. 3. The lung image in axial plane

*B. Gray Record*

Before start the enhancement step, gray level in the image input will be recorded. Gray level that recorded is gray color in some spot inside the cancer area. This step propose to make the template of cancer color, it is useful for evaluating non-cancer area in the next step. The gray template are store in txt file. We are using color picker with mouse clicker to get the cancer color. First record are stored in the txt file directly. The next record are checked before stored to overcome the similarity of color. The procedure are shown in Fig. 4.

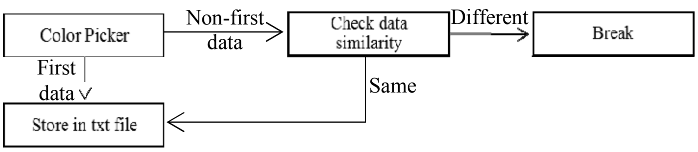
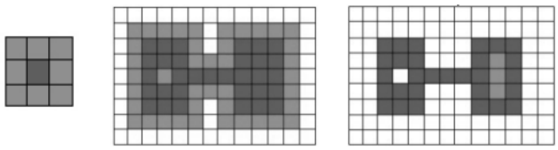



Fig. 4. Schematic diagram of gray record

*C. Erode and Dilate*

This step is using erode function to eliminate noise. Step continue with dilate function to isolate of individual elements and link disparate elements. Both dilation and erosion are produced by the interaction of a set of structure elements with the image. This structure element has the shape and origin shown in Figure 5(a). Suppose A is the set of pixels, and B is a structure element. Structure element B dilates set A by filling holes with certain shapes and sizes provided by the structure elements, such as in Figure 5(b). Structure element B erodes set A by eliminating certain shapes and sizes of the structure provided by the structure elements, such in as Figure 5(c) [10].



(a) (b) (c)

Fig. 5. Morphology operations (a) structure element (b) dilation (c) erosion

The element was set with an elliptic structuring in size 7, 7. The output of this step is image with smoother texture as an input for watershed transformation.

*D. Watershed*

Watershed transformation is used to segment lung cavity area. Watershed applied because lung cavity area are different in each image, but it still has the same center point. Input image consists of many segments, although the cavity area has less complicated segments. First step of watershed transformation is mark particular area that will be segmented. This study uses 3 marks on particular areas. Areas consists of non-cavity area (outer layer of body, bones, skin, pleura, etc.), right cavity area and left cavity area. Non-cavity area is marked by a vertical line positioning from (0, 0) to (0, maximum height of image), therefore vertical mark will cut pleura area, skin and bones. While the other 2 marks used dot mark. Where non-cavity area as the waters and cavity area as the land. Thus it will be split into two segments, cavity and non-cavity.

*E. Crop*

Subsequently, image will be crop in segmented area (land area in watershed). The output of this step is cavity area (without pleura). The procedure are shown in Fig. 5.

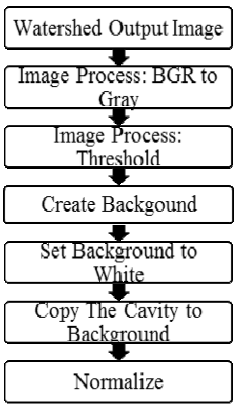



Fig. 6. Cropping procedure

### *F. Color Histogram Calculation*

The output of the watershed will be calculated and mapped by histogram calculation. The image input of histogram calculation is a gray image that requires only one curve line. Size of the histogram is provided at 256 levels, which the gray values ranging from 1 to 255. The zero level is not calculated because the gray with 0 value is air area. The histogram graph is depicted on a plane with 512 x 400 pixels size. This step purpose to find. This step aims to collect gray data on the image. Collected data can be useful for describing the gray level of the image and useful for the calculation mean step.

### *G. Mean and Standard Deviation*

The mean and standard deviation calculations are useful for determining the range of gray values that will be segmented. Gray values in the range will be maintained, in this case the cancer area. First, the images will be scanned by matching the gray templates. The mean and standard deviation calculations are useful for determining the range of gray values that will be segmented. Gray values in the range will be maintained, in this case the cancer area. First, the images will be scanned by matching the gray templates. Appropriate values will be collected and calculated. The procedure are shown in Fig. 6.

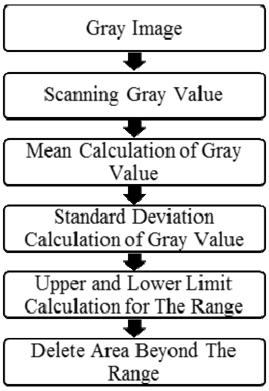



Fig. 7. Schematic diagram of segmentation

Range is value between the upper and lower limits. The upper limit value is derived from the mean plus the standard deviation value, while the lower limit value is derived from the mean value minus the standard deviation value. Fig. 6 is an illustration of the normal mean and the standard deviation curve [9].

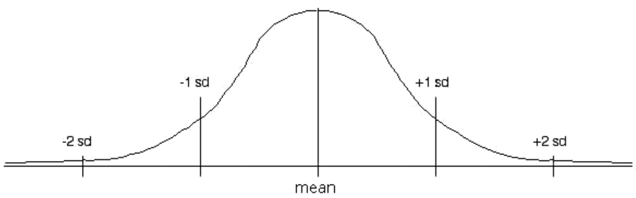



Fig. 8. Illustration of the normal mean and the standard deviation curve

### *H. Binary Image*

Result image of cutting process will be converted into binary image inverted. This step will make the entire cavity segment has white color (255, 255, 255) and the other has black color (0, 0, 0). This will make it easier to calculate segmented area based on pixels. In this study, we used threshold binary inverted with minimum value is 140 and maximum value is 255.

### *I. Median Filter*

This study used median filter to get smooth texture and reduce noise after segmentation process. Filter applied with kernel size of 19. Output image from binary step will be processed by median filter to get cavity area. This step will produce area to crop original image and result cavity image.

### *J. Area and Percentage Calculation*

There are two areas will be counted, cavity area and cancer area. Both has pixel-based calculation. In this study, authors used non-zero pixel calculation. This calculation will be converted in millimeter cubic. Conversion drawn from sum of pixels multiply with constant. The formulation of area calculation is shown in formula (1).

$$Area = \left( \sum_{i=max}^{i=0} pixel \right) x\, 0.26 \qquad (1)$$

Percentage calculation is done by comparing two calculated areas. Then will be obtained percentage of cancer blockage in lung cavity. The formulation is shown in formula (2).

$$\%\ Blockage = \frac{Cancer\ area}{Cavity\ area} x\, 100\% \qquad (2)$$

## III. Contribution

Contribution of this paper is application of automatic lung image segmentation for cancer and lung cavity segmentation to calculate segmented area and determine blockage percentage of cancer inside lung cavity.

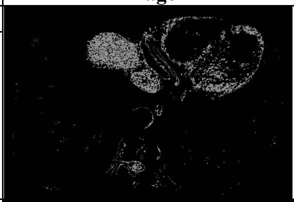
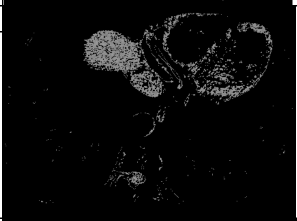
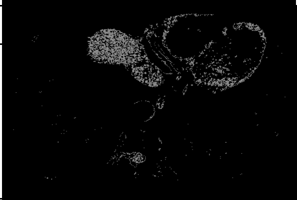
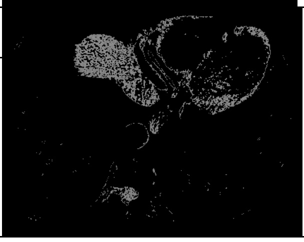
## IV. Experiment Result

The proposed algorithm was tested using dataset axial pulmonary CT-Scan image from Hospital Haji Surabaya. The experimental steps consist of gray record, cavity segmentation, cancer segmentation, percentage area calculation. Segmentation and calculation testing were performed on 251 axial pulmonary CT-Scan images.

### *A. Gray Record*

Training data by recording gray values at some point on the cancer object. This step will generate the template. The test will determine the effect of number of points on the image segmentation result. Table 1. shows the effect of number of gray.

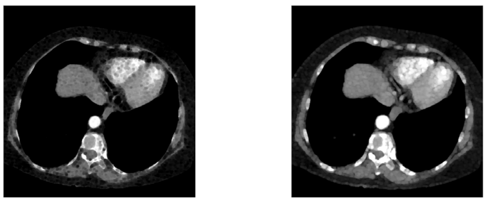
Table 1. The effect of number of gray

| No. | Gray | Image |
|---|---|---|
| 1. | rows: 15<br>data:<br>[ 123, 158, 161, 122, 159, 144, 157, 167, 153, 135, 126, 128, 118, 151, 129 ] |  |
| 2. | rows: 10<br>data:<br>[ 136, 132, 111, 139, 148, 147, 149, 119, 174, 128 ] |  |
| 3. | rows: 5<br>data:<br>[ 133, 123, 141, 161, 137 ] |  |
| 4. | rows: 3<br>data:<br>[ 155, 132, 112 ] |  |

Testing results the decision that template has 3 points gay value with data values: [155, 132, 112]. The number of 10 points also has corresponding segmentation results, but many points would be an obstacle. The program will run longer because many gray values are scanned. The number of 3 points with data values: [155, 132, 112] is the optimal values for segmentation template.

## B. *Cavity Segmentation*

First, image noise will be removed by erosion and dilation. Elements for the erosion and dilation process are ellipse with size 6.6. Examples of erosion result are shown in the Fig. 8a and example of dilation result are shown in the Fig. 8b.



a b

Fig. 9. Result (a) erosion (b) dilation

Watershed transformation begins by marking 3 areas on the output image of previous process. The area includes the outer of body area, right cavity area and left cavity area. Three marks representation are shown in the Fig. 9.


Fig. 10. Watershed mark

This process will segment the cavity area and eliminate the non-cavity area. The sample of watershed transformation result are shown in the Fig. 10.

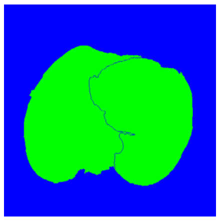


Fig. 11. Watershed segment

There are two colors shown in Fig. 10. Green color are cavity area and blue color are non-cavity area. This output image will be cropped as in Fig. 11.

Fig. 12. Crop result

The watershed image results will be converted into binary image before calculated. Binary image of cavity image are shown in Fig. 12.

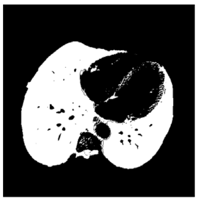


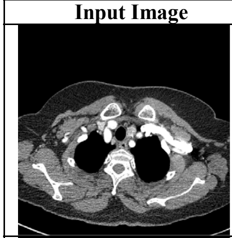
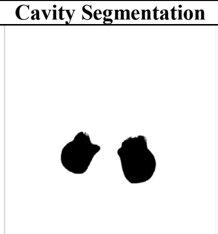
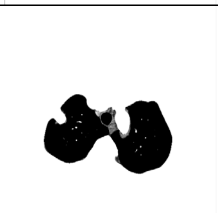
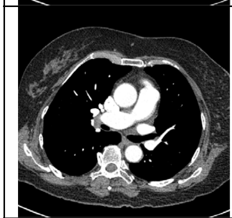
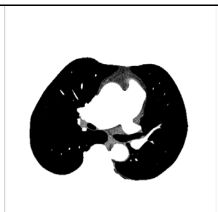
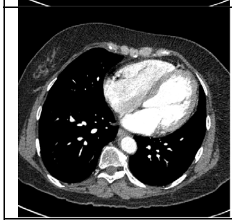
Fig. 13. Binary image of cavity image

Binary image will be filtered by median filter to get smooth texture and reduce noise. The result of median filter will be used as a reference for cutting the cavity area again. The second cutting purpose to produce clean lung image (without heart). Result for second cutting after median filter are shown in Fig. 13.

Fig. 14. Second crop after median filter

Cavity area will be calculated based on non-zero pixels. Then sum of non-zero pixels will be converted into millimeter by multiply with constant. Some results of cavity segmentation and area calculation are shown in table 2.

Table 2. Cavity segmentation and calculation

| Input Image | Cavity Segmentation | Calculation ($mm^2$) |
|---|---|---|
|  |  | 3641.56 |
|  |  | 9288.50 |
|  |  | 18457.1 |
|  |  | 18688.2 |

From Table 2. we know that experiment has some limitations. When cancer attached with pleura, it would not be detected, because cancer will be read as pleura. Error will be calculate with this formula (3).

$$\%\ error = \left|\frac{Theoretical - Experimetal}{Theoretical}\right| x\,100 \qquad (3)$$

The experiment have been tested and generate 5.92% error value. Table 3. shows some error value calculations.

Table 3. Error value calculations in the cavity segmentation

| Experimental ($mm^2$) | Theoretical ($mm^2$) | Error (%) |
|---|---|---|
| 4873.18 | 5279.04 | 7.68814 |
| 7671.04 | 8609.12 | 10.89635 |
| 10933 | 12032.28 | 9.136091 |
| 11697.92 | 13633.88 | 14.19963 |
| 12013.3 | 15670.46 | 23.33792 |
| 20416.5 | 15953.6 | 27.97425 |
| 19966.18 | 17491.24 | 14.1496 |
| 16834.22 | 17350.06 | 2.973131 |
| 21047.52 | 17709.38 | 18.84956 |
| 19653.4 | 18403.32 | 6.792687 |
| 5038.28 | 5157.62 | 2.313858 |
| 14790.88 | 11781.12 | 25.54732 |
| 15719.34 | 15093.26 | 4.148077 |
| 19550.44 | 17163.12 | 13.90959 |
| 22441.64 | 17785.82 | 26.17715 |
| 27031.42 | 19735.3 | 36.9699 |
| 22564.62 | 20933.9 | 7.789853 |
| 20723.56 | 21727.42 | 4.620245 |
| 21498.1 | 21396.18 | 0.476347 |
| 3641.56 | 5117.84 | 28.84576 |
| 8411.78 | 9403.42 | 10.54552 |
| 13574.34 | 12700.22 | 6.882715 |
| 16647.8 | 16456.18 | 1.164426 |
| 17206.02 | 17771.78 | 3.183474 |
| 18513.56 | 18993.78 | 2.528301 |
| 19775.86 | 20786.74 | 4.8631 |
| 18403.32 | 20800 | 11.5225 |
| 22041.5 | 20152.08 | 9.375806 |
| 19665.88 | 21871.98 | 10.08642 |
| 4411.68 | 5944.64 | 25.78726 |
| 8603.14 | 10298.08 | 16.4588 |
| 9966.06 | 14584.7 | 31.66771 |
| 12184.12 | 17662.32 | 31.01631 |
| 11146.2 | 18733 | 40.49965 |
| 12298.26 | 21279.18 | 42.2052 |
| 19206.72 | 21801.26 | 11.90087 |
| 21251.62 | 23152.48 | 8.210179 |
| 22222.72 | 23687.3 | 6.182976 |
| 27069.64 | 23620.22 | 14.60367 |
| 4613.44 | 3178.76 | 45.13332 |
| 7581.6 | 7819.5 | 3.042394 |
| 10167.3 | 11903.06 | 14.58247 |
| 12635.48 | 14977.3 | 15.6358 |
| 16537.82 | 17079.4 | 3.170954 |
| 17207.06 | 17534.4 | 1.866845 |
| 16754.14 | 17412.98 | 3.783614 |
| 14652.82 | 16191.76 | 9.504464 |
| 14013.22 | 15262.78 | 8.186975 |
| 13299.52 | 14761.24 | 9.90242 |
| 4914.26 | 6174.22 | 20.40679 |
| 8949.98 | 9545.64 | 6.240126 |
| 13225.42 | 13874.38 | 4.677398 |
| 15716.48 | 16221.14 | 3.111125 |
| 17863.56 | 18990.14 | 5.932447 |
| 16688.36 | 19828.9 | 15.8382 |
| 16673.54 | 20311.72 | 17.91173 |
| 19150.04 | 20681.7 | 7.405871 |
| 28033.2 | 20431.84 | 37.2035 |
| 21840 | 19816.68 | 10.21019 |
| 3799.64 | 4398.42 | 13.61352 |
| 8176.74 | 8786.18 | 6.936348 |
| 12380.94 | 12786.02 | 3.168148 |
| 14517.36 | 14835.34 | 2.143395 |
| 15106.78 | 17531.28 | 13.82957 |
| 14868.62 | 18556.98 | 19.87586 |
| 15529.8 | 19472.7 | 20.24835 |
| 17399.98 | 19966.44 | 12.85387 |

| | | |
|---|---|---|
| 17861.48 | 19325.02 | 7.573291 |
| 17718.48 | 19282.9 | 8.112991 |
| 4947.02 | 5679.7 | 12.89998 |
| 8580.78 | 9731.54 | 11.82506 |
| 12465.96 | 13756.6 | 9.381969 |
| 15131.22 | 15537.34 | 2.613832 |
| 17219.02 | 17765.54 | 3.076293 |
| 17155.32 | 18304.78 | 6.279562 |
| 16903.64 | 18825.04 | 10.20662 |
| 17858.36 | 19696.56 | 9.332594 |
| 16702.4 | 18134.48 | 7.897001 |
| 16907.54 | 18163.34 | 6.913927 |
| **Average** | | 12.75961 |

## *C. Cancer Segmantation*

Cancer segmentation input is taken from watershed transformation output. Image will be cropped and processed with color histogram calculation. Image will be scanned based on the template of gray record step. Recorded gray values will be calculated to find mean and standard deviation. Then we get upper limitation range and lower limitation ranger. These range will evaluate cancer object and produce a cancer image. Fig. 8. shows an example of the segmentation result.

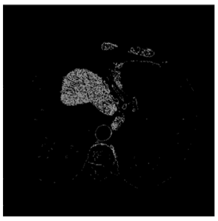
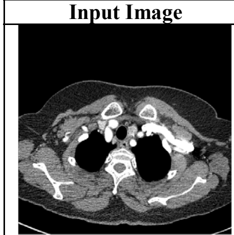
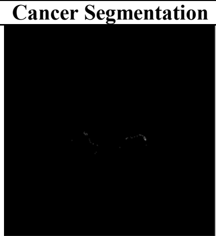
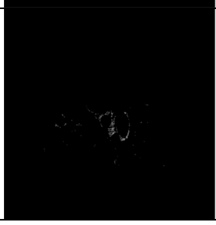
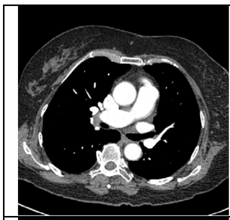
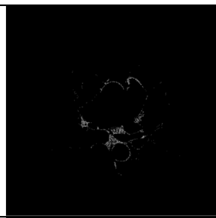
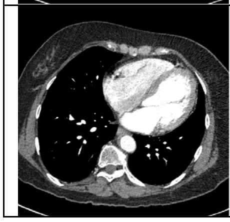
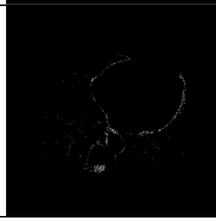


Fig. 15. Cancer segmentation result

Cancer area will be calculated based on non-zero pixels. Then sum of non-zero pixels will be converted into millimeter by multiply with constant. Some results of cancer segmentation and cancer area calculation are shown in Table 4.

Table 4. Cancer segmentation and calculation

| Input Image | Cancer Segmentation | Calculation ($mm^2$) |
|---|---|---|
|  |  | 19.76 |
|  |  | 152.10 |
|  |  | 426.66 |
|  |  | 388.18 |

From Table 2. we know that experiment has some limitations. Heart covering layer is still detected, thus it will read as cancer object because it has similar gray values. The experiment have been tested and generate 35.82% error value. Table 5. shows some error value calculations.

Table 5. Error value calculations in the cancer segmentation

| Experimental ($mm^2$) | Theoretical ($mm^2$) | Error (%) |
|---|---|---|
| 874.9 | 1039.48 | 15.83292 |
| 230.36 | 390.78 | 41.05123 |
| 300.3 | 374.14 | 19.73593 |
| 4384.12 | 3253.38 | 34.75585 |
| 3738.02 | 4424.16 | 15.50893 |
| 5571.28 | 4363.32 | 27.68442 |
| 5101.2 | 5224.18 | 2.354054 |
| 1619.28 | 1803.62 | 10.22056 |
| 2314.26 | 2826.2 | 18.11408 |
| 2134.34 | 1966.64 | 8.527234 |
| 1061.32 | 2600.78 | 59.19224 |
| 1981.98 | 4472.78 | 55.68796 |
| 2838.68 | 6604.78 | 57.02082 |
| 2190.76 | 4884.88 | 55.15222 |
| 829.66 | 1375.92 | 39.70144 |
| 140.92 | 127.14 | 10.83845 |
| 113.88 | 134.68 | 15.44402 |
| 164.06 | 113.62 | 44.39359 |
| 312.78 | 437.58 | 28.5205 |
| 389.48 | 796.9 | 51.12561 |
| 753.48 | 1701.18 | 55.70839 |
| **Average** | | 31.74145 |

# V. Conclusion

The proposed method presents solution for segmentation of lung image for cancer disease. This study has produced automatic lung cancer detection that is follow with area calculation and percentage of blockage calculation. The cavity segmentation using watershed transformation has less error value. The cancer segmentation using color histogram has some errors if heart cover layer is segmented. The performance segmentation for assessment errors lung obtained an average cavity area segmentation 12.75% and cancer area segmentation 31.74%.

## Acknowledgment

The authors would like to thank the Hospital Haji Surabaya for giving input datasets and also would like to thank the Politeknik Elektronika Negeri Surabaya.

## References

[1] P.-y. Yu, J.-w. Yan, G.-p. Zhang and M.-s. Lio, "The Application of the Watershed Algorithm Based on Line-Encoded in Lung CT Image Segmentation," Intelligent Systems and Applications (ISA), May 2011.

[2] J. Wang, H. Lu, G. Eude and Q. Lu, "A Fast Region Merging Algorithm For Watershed Segmentation," ICSP '04 Proceedings, vol. 1, pp. 781 - 784, September 2004.

[3] H. Sun, J. Yang and M. Ren, "A fast watershed algorithm based on chain code and its application in image segmentation," vol. 26, no. 9, p. 1266 – 1274, Juli 2005.

[4] A. S, Fisika Radiologi, Surabaya, Jawa Timur: Radiologi FK UWKS.

[5] K. M Schramm, MD, "Acute Pulmonary Embolism (Helical CT)," Drugs, Diseases & Procedures, 6 Maret 2016.

[6] S. N. E. Kostas Haris, N. Maglaveras and A. K. Katsaggelos, "Hybrid Image Segmentation Using Watersheds and Fast Region Merging," Transactions on Image Processing, vol. 7, no. 12, pp. 1684 - 1699, 6 Agustus 2002.

[7] S. Guo and L. Wang, "Automatic CT Image Segmentation of the Lungs with An Iterative Chan-Vase Algorithm," Informatics, Electronics & Vision (ICIEV), June 2015.

[8] D. S. Sianny Suryawati, Emergency Radiology: The Basics, Surabaya: Radiology Department, Faculty of Medicine, Wijaya Kusuma University, 2016.

[9] Sugiyono, Metode Penelitian Bisnis (Pendekatan Kuantitatif, Kualitatif, dan R&D), Bandung: Alfabeta, 2009.

[10] SIGIT, Riyanto; ARIEF, Zainal; BACHTIAR, Mochamad Mobed. Development of Healthcare Kiosk for Checking Heart Health. EMITTER International Journal of Engineering Technology, [S.l.], v. 3, n. 2, apr. 2016. ISSN 2443-1168. Available at: <https://emitter.pens.ac.id/index.php/emitter/article/view/49/47>. Date accessed: 29 aug. 2017. doi:http://dx.doi.org/10.24003/emitter.v3i2.49.