PAPER • OPEN ACCESS

Automatic Solitary Lung Nodule Detection in Computed Tomography Images Slices

IOP Conf. Series: Journal of Physics: Conf. Series **953** (2018) 012068 doi:10.1088/1742-6596/953/1/012068

Automatic Solitary Lung Nodule Detection in Computed Tomography Images Slices

I W B Sentana[1], N Jawas[2], S A Asri[1]

[1]Politeknik Negeri Bali, Jl. Raya Bukit Jimbaran, Bali, Indonesia
[2]STMIK STIKOM Bali, Jl. Raya Puputan No. 86, Denpasar, Indonesia
e-mail : budisentana@pnb.ac.id

Abstract. Lung nodule is an early indicator of some lung diseases, including lung cancer. In Computed Tomography (CT) based image, nodule is known as a shape that appears brighter than lung surrounding. This research aim to develop an application that automatically detect lung nodule in CT images. There are some steps in algorithm such as image acquisition and conversion, image binarization, lung segmentation, blob detection, and classification. Data acquisition is a step to taking image slice by slice from the original *.dicom format and then each image slices is converted into *.tif image format. Binarization that tailoring Otsu algorithm, than separated the background and foreground part of each image slices. After removing the background part, the next step is to segment part of the lung only so the nodule can localized easier. Once again Otsu algorithm is use to detect nodule blob in localized lung area. The final step is tailoring Support Vector Machine (SVM) to classify the nodule. The application has succeed detecting near round nodule with a certain threshold of size. Those detecting result shows drawback in part of thresholding size and shape of nodule that need to enhance in the next part of the research. The algorithm also cannot detect nodule that attached to wall and Lung Chanel, since it depend the searching only on colour differences.

1. Introduction

Lung Nodule or Pulmonary Nodule are white patches resembling cotton or clouds that are sometimes seen in medical images scanned on thorax. Lung Nodule is used as a marker or early symptom for the certain lung disease, such as Bronchitis, Cystic fibrosis, Emphysema, Pneumonia, Tuberculosis, pulmonary edema or even lung cancer [1]. Based on World Cancer Report data of 2014, lung cancer remains one of the deadliest diseases in the world, where every year, there are 1.2-1.56 million deaths caused by this disease [2], [3]. The number will continue to increase, which in 2030 is predict to occur 17 million deaths from lung cancer [4]. From the research conducted by [5], 80% of new lung cancers are known after suffering from middle to upper stage disease. The existence of early detection based on lung nodule will be able to provide better handling to the patient.

Nodule can be detected by utilizing various techniques in the radiology such as the use of X-ray (Rontgen), Computed Tomography (CT) or using Magnetic Resonance Imager (MRI) [4], [6], [7]. The image generated by the Rontgen process is only best used in the early screening stage of the possibility of nodule, whereas to know the location and size of the nodule more accurately, it is usually used CT scan technique or MRI[1]. MRI does provide more detailed image results, but MRI has a weakness that is expensive and not very good on soft tissue, as well as lungs. Those makes the CT scans are still in great demand for use and research. Although it gives better results than X-ray images, detecting nodules automatically is still challenging.

Automated application of lung nodule detection will greatly assist radiologist in alleviating its work. Based on research conducted by (Orozco et al., 2013), each radiologist takes an average of 15-20 minutes to detect nodules manually. On the same day every radiologist performs 45 tiring examinations on medical images. Hence, the number of research to detect lung nodule automatically by utilizing image processing techniques increase rapidly.

2. Research Methodology

This research consist of few steps, including data acquisition and conversion, image binarization, lung segmentation, nodule candidate detection and nodule classification as shown in Figure 1. Meanwhile the CT images data obtained from Data Science Bowl repository, especially for lung CT images.

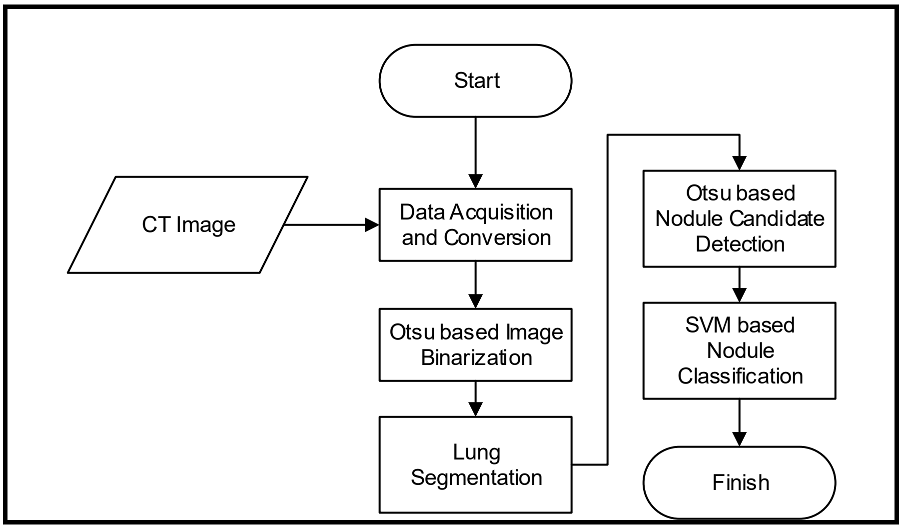

Figure 1. Research Steps.

2.1. Data Acquisition and Conversion

The first step of algorithm is taking the image slice by slice. The original format of CT images is in *.dicom that can consist of thousand slices. Each slices is in the form of two dimensional image that consist of certain pixel, unlike the original form which is known as voxel. In order to analyzed the images, each slices of the CT image need to convert to another format, which in this case will be convert into *.tif. the metadata of each slices is also saved and adjusted to the new format.

2.2. Image Binarization

After all slices converted, then the next step is separate the foreground and background of the images. This step is need to detect the part of the body only and removed rest of image part that taken during photo process. The point of this steps in increasing contrast between the darker and brighter part of the images. Hence, Otsu algorithm is used to assist this step.

Supposed that the pixel in given image be represented in L gray level *(1,2,3….,L)*. Let n_i denote the number of pixel at level i and N denote the total number of Pixel, $N = \sum_i^L n_i$. The probability of occurrence of level i is given by $pi = n_i/N$. Let an image divided into two classes C_0 and C_1, by threshold $T.C_0$ consist of pixels with level *[1,….,T]* and C_1 consist of pixels with level *[T+1,…., L]*. Let $P_0(T)$ and $P_1(T)$ denoted the cumulative probabilities, $\mu_0(T)$ and $\mu_1(T)$ denote the mean level, and, $\sigma_0^2(T)$ and $\sigma_1^2(T)$ denote the variances of the classes C0 and C1, respectively. This value are given by:

$$P_0(T) = \sum_{i=1}^{T} p_i \tag{1}$$

$$P_1(T) = \sum_{i=T+1}^{L} p_i = 1 - P_0(T) \tag{2}$$

$$\mu_0(T) = \sum_{i=1}^{T} i \frac{p_i}{P_0(T)} = \frac{1}{P_0(T)} \sum_{i=1}^{T} ip_i \tag{3}$$

$$\mu_1(T) = \sum_{i=T+1}^{L} i \frac{p_i}{P_1(T)} = \frac{1}{P_1(T)} \sum_{i=T+1}^{L} ip_i \tag{4}$$

$$\sigma_0^2(T) = \sum_{i=1}^{T} (i - \mu_0(T))^2 \frac{p_i}{p_0(T)} \tag{5}$$

$$\sigma_1^2(T) = \sum_{i=T+1}^{L} (i - \mu_1(T))^2 \frac{p_i}{p_1(T)} \tag{6}$$

Let μ , $\sigma_b^2(T)$, and $\sigma_w^2(T)$ represent the mean level of the image, between-class variance and the within-class variance, respectively:

$$\mu = \sum_{i=1}^{L} ip_i = P_0(T)\mu_0(T) + P_1(T)\mu_1(T) \tag{7}$$

$$\sigma_b^2(T) = P_0(T)(\mu_0(T) - \mu)^2 + P_1(T)(\mu_1(T) - \mu)^2 \tag{8}$$

$$\sigma_w^2(T) = P_0(T)\sigma_0^2(T) + P_1(T)\sigma_1^2(T) \tag{9}$$

The threshold decided by maximizing the between-class variance proposed in Otsu is:

$$T^* = \underset{1 \le T < L}{\arg\max} \left\{\sigma_b^2(T)\right\} \tag{10}$$

This value is equal to the threshold decided by minimizing the within-class variance criterion:

$$T^* = \underset{1 \le T < L}{\arg\min} \left\{\sigma_w^2(T)\right\} \tag{11}$$

Furthermore, the above threshold is same as the threshold calculated by maximizing the ratio between-class variance to within-class variance [8]

2.3. Lung Segmentation

After separate between foreground and background of CT image, then the next step is removing the background part of image. More contrast foreground images then taken and by using closest value of pixel, surrounding identical pixel value are grouped to find certain shape. This step than will separated the Lung part only. Moore Neighborhood Tracing (MNT) algorithm is tailoring to assist this process. MNT is pixel-following based algorithm that traces contour pixels in a predefined manner and then saves their coordinates in memory according to the trace order. This algorithm finds the next contour pixel using eight connected chain codes with a clockwise sequence starting from the rear pixel of the tracer, i.e., the tracer first moves toward the rear *(T(P_{Rear},d_{Rear}))* and finds the next clockwise contour pixel, such as the left-rear, left, font-left, front, front-right, right and rear-right pixels[9].

2.4. Nodule Candidate Detection

Lung shape that acquired from segmentation process than isolated to localized the nodule candidate only on that area. In this step, again Otsu algorithm proposed by Nobuyuki Otsu in 1979[10], was use to enhance the differences between lung area and nodule that has brighter color.

2.5. Nodule Classification

The final step of this algorithm is tailoring Support Vector Machine (SVM) to classify the nodule candidate, whether its nodule or not. SVM consist of two main steps which is training and classification itself. SVM is based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. For the linearly separable samples, the optimal classification hyperplane can separate the instances into two categories. For the linearly inseparable problems, the instances in the original space will be mapped into the high-dimensional feature space by using a nonlinearly transformation. To construct an optimal hyperplane, SVM employs an iterative training algorithm, which is used to minimize an error function. A general SVM classification can be described as a mathematical optimization problem[11][12][13][14]:

$$\min \Phi(w) = \frac{1}{2}\left\|w\right\|^2 + C\sum_{i=1}^{N}\xi_i$$

$$\text{s.t } y_i(w^T.\phi(x_i)+b) \geq 1-\xi_i, (i=1,2,3,...N) \tag{12}$$

where C is the penalty parameter, and ξ_i represents parameter for handling inseparable data. Usually, minimizing the object function $\Phi(w)$ of formula (12) needs:(1) maximizing the margin between two classes; (2) minimizing the misclassifying rate. The parameter C controls the trade-off between the slack variable penalty and the size of the margin. The index i labels the N training cases, $y_i \in \pm 1$ represents the class labels and x_i represents the independent variables. The kernel function ϕ is used to transform data from the input to the feature space.

The kernel function ϕ is used to transform data from the input to the feature space. Three kinds of kernel functions called Linear kernel, Polynomial kernel and RBF kernel are commonly used. In this paper, we choose the RBF kernel as the kernel function of SVM, which is shown in formula (13).

$$k(x_i, x_j) = \exp(-\gamma\left\|x_i - x_j\right\|^2) \tag{13}$$

Where γ is inversely proportional to the width of the kernel. Traditionally, SVM classifier uses a defaults set of C and γ in solving the pattern classification problems[15].

3. Result and Discussion

The algorithm is implemented successfully according those step above. Figure 2 until Figure 6 showing the result of each step within algorithm. Figure 2 showing converted CT image in *.tif format. It is supposed to have big nodule on the left side attached to bronchus.

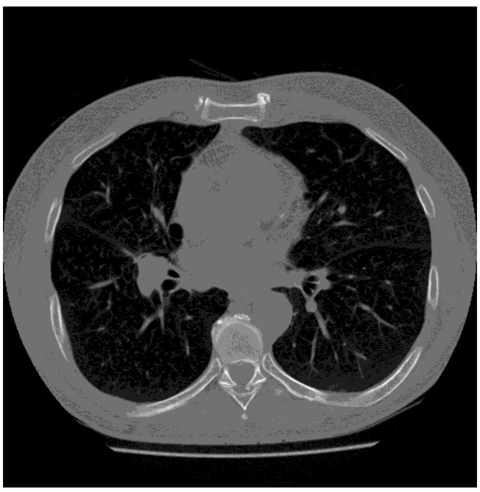

Figure 2. Lung CT Image after acquisition and conversion

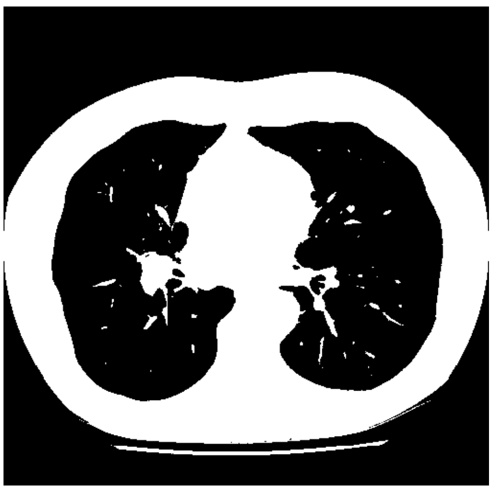

Figure 3. Image after Otsu Based Binarization

Meanwhile Figure 3 showing the result after Otsu binarized the image. Its shows more contrast on those area. The goal is to define the threshold of area so the lung area can separated from whole image. Unfortunately, the nodule on the left lung, that supposed to be part of Lung, is exclude and detected as another soft tissue, since its color similarity.

Figure 4 showing the image after segmentation process according to MNT algorithm, where all edge of lung is already detected based on threshold in Otsu algorithm tailored previously. More contrast color between lung and nodule are shown in Figure 5 after Otsu algorithm is use just in lung area. The white blob is nodule candidate that can be real nodule or can be a part of lung.

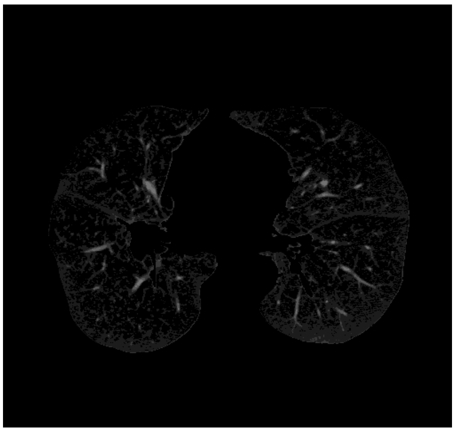

Figure 4. Lung image after Segmentation

Figure 5. After Tailoring Otsu to separate between bright and dark part of image

Those nodule candidate than classify using SVM to detect whether it nodule or not. The result is shown in Figure 6. It is shows that some near round nodule that have more than 20 pixel in size was detected. The threshold of near round and number of pixel is setup manually. It is also shown that big nodule on left part Lung attached to bronchus is exclude as a nodule.

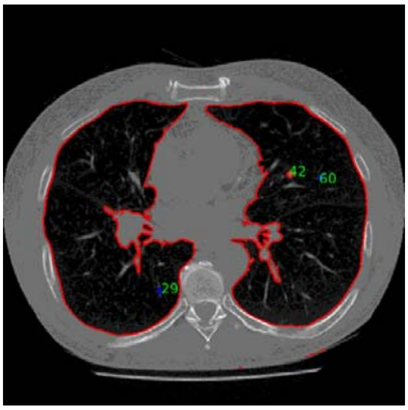

Figure 6. After involving SVM to classify nodule. Nodule coordinate is reverse to previous step on original images. Some near round nodule is detected by the application

The grown truth analysis that have already conduct in Tabanan Regency General Hospital is also giving some advice to detect non near round nodule, since not all nodule are in that shape. Application testing using standard data is also needed to compare to the others algorithm to measure the accuracy of the algorithm. Enhancement on pre-processing data will become future research to get better result on classification, especially to detect nodule that attached to Lung Wall or Chanel.

4. Conclusion

The algorithm is implemented successfully according to the step in Figure 1. It also successfully detect nodule that have near round nodule that have more than 20 pixel size. However the algorithm still have some drawback, especially in detecting nodule that attached to the wall or Lung Chanel. Some enhancement on segmentation process need to improve to get a better result on classification. The algorithm also need to test using standardized lung nodule dataset, so the algorithm can be comparable to the others.

5. Acknowledgments

Authors wish to acknowledge that this research is funded by Directorate General of Indonesia Higher Education, Ministry of Higher Education and Research and Technology, Republic of Indonesia. This research also supported by Public Hospital of Tabanan Regency as corresponding Grown Truth testing.

6. References

[1] N. Mesanovic, M. Grgic, H. Huseinagic, M. Males, E. Skejic, and M. Smajlovic, "Automatic CT Image Segmentation of the Lungs with Region Growing Algorithm," *Int. J. Eng. Adv. Technol.*, vol. 4, no. 5, pp. 395–400, 2015.

[2] H. M. Orozco, S. Member, O. Osiris, V. Villegas, S. Member, and H. De Jes, "Lung Nodule Classification in CT Thorax Images using Support Vector Machines," in *12th Mexican International Conference on Artificial Intelligence*, 2013, pp. 277–282.

[3] P. Bhuvaneswari and A. B. Therese, "Detection of Cancer in Lung With K-NN Classification Using Genetic Algorithm," *Procedia Mater. Sci.*, vol. 10, no. Cnt 2014, pp. 433–440, 2015.

[4] K. Mya, M. Tun, and A. S. Khaing, "Feature Extraction and Classification of Lung," *Int. J. Eng. Res. Technol.*, vol. 3, no. 3, pp. 2204–2210, 2014.

[5] F. V. Farahani, A. Ahmadi, and M. H. F. Zarandi, "Lung Nodule Diagnosis from CT Images Based on Ensemble Learning," 2015.

[6] X. Gao, C. Chu, Y. Li, P. Lu, W. Wang, W. Liu, and L. Yu, "The method and efficacy of support vector machine classifiers based on texture features and multi-resolution histogram from 18 F-FDG PET-CT images for the evaluation of mediastinal lymph nodes in patients with lung cancer ଝ," *Eur. J. Radiol.*, vol. 84, no. 2, pp. 312–317, 2015.

[7] C. Bhuvaneswari, P. Aruna, and D. Loganathan, "Classification of Lung Diseases by Image Processing Techniques Using Computed Tomography Images," *Int. J. Adv. Comput. Res.*, vol. 4, no. 1, pp. 2277–7970, 2014.

[8] X. Xu, S. Xu, L. Jin, and E. Song, "Characteristic analysis of Otsu threshold and its applications," *Pattern Recognit. Lett.*, vol. 32, no. 7, pp. 956–961, 2011.

[9] J. Seo, S. Chae, J. Shim, D. Kim, C. Cheong, and T. Han, "Fast Contour-Tracing Algorithm Based on a Pixel-Following Method for Image Sensors," *MDPI J. sensors*, vol. 16, no. 353, pp. 1–27, 2016.

[10] C. Sha, J. Hou, and H. Cui, "A Robust 2D Otsu ' s Thresholding Method in Image Segmentation," *J. Vis. Commun. Image Represent.*, no. October, 2016.

[11] P. Chen, L. Yuan, Y. He, and S. Luo, "An improved SVM classifier based on double application in analogue circuit diagnosis Reference : To appear in : Neurocomputing," *Neurocomputing*, 2015.

[12] H. R. Tizhoosh and F. Khalvati, "Medical Image Classification via SVM using LBP Features from Saliency-Based Folded Data," pp. 0–4, 2015.

[13] P. Eskandarian and J. Bagherzadeh, "Computer-Aided Detection of Pulmonary Nodules based on SVM in Thoracic CT Images," in *7th International Conference on Information and Knowledge Technology*, 2015, pp. 0–5.

[14] C. Désir, C. Petitjean, L. Heutte, L. Thiberville, and M. Salaün, "Computerized Medical Imaging and Graphics An SVM-based distal lung image classification using texture descriptors," *Comput. Med. Imaging Graph.*, vol. 36, no. 4, pp. 264–270, 2012.

[15] M. Keshani, Z. Azimifar, F. Tajeripour, and R. Boostani, "Lung nodule segmentation and recognition using SVM classifier and active contour modeling: A complete intelligent system," *Comput. Biol. Med.*, vol. 43, no. 4, pp. 287–300, 2013.