# Sensitivity improvement of automatic pulmonary nodules detection in chest X-ray CT images

Satoshi Shimoyama[1], Noriyasu Homma[2], Tadashi Ishibashi[1] and Makoto Yoshizawa[2]

[1] *Tohoku University Graduate School of Medicine, 2-1 Seiryo-machi, Aoba-ku, Sendai, 980-8575 JAPAN*
[2] *Cyberscience Center, Tohoku University*
[2] *(Tel : 81-022-795-7130; Fax :81-022-263-9163)*
*(E-mail: simoyama@yoshizawa.ecei.tohoku.ac.jp)*

***Abstract***: In this paper, we develop an automatic detection method of non-isolated pulmonary nodules as a part of computer-aided diagnosis (CAD) system for lung cancers in chest X-ray CT images. An essential core of the method is to separate non-isolated nodules from connecting structures like chest wall and blood vessels. The isolated nodules can be detected more easily by CAD syetems developed previously. To this end, we propose a binarization technique by using two thresholds as a preprocessing for nodule candidates. We evaluate the performance using the receiver operating characteristic (ROC) analysis in clinical chest CT images. The results suggest that detection ability of non-isolated nodules by the proposed method is superior to that by the conventional preprocessing methods.

***Keywords***: computer-aided diagnosis, lung nodule, X-ray CT, lung cancer, multiple thresholds.

## I. INTRODUCTION

It is very important to detect cancers at an early stage. To this end, the X-ray computed tomography (CT) should be used because pulmonary nodules that are typical shadows of pathological changes of lung cancers and can be depicted more clearly compared to use of the chest X-ray examination [1-3]. This is true even the nodules are at early stages. A problem to use the CT is that radiologists have to diagnose at least more than 30 image slices per patient. Thus, some computer-aided diagnosis (CAD) systems have been proposed to help their diagnosis work. Some excellent results have been reported for isolated pulmonary nodules [4-6]. The previous methods are however inaccurate for non-isolated nodules connected to the chest walls and blood vessels, etc.

In this paper, as a preprocessing, we propose a novel binarization method for detection of non-isolated nodules. The core of this method is to use two thresholds in order to convert such non-isolated nodules into isolated ones. The effectiveness of the proposed method is verified by using to some clinical cases of lung cancer CT images.

## II. AUTOMATIC NODULES DETECTION

### 1. Variable N-Quoit filter

The CAD system needs to detect suspicious lesions. It is used the variable N-Quoit filter for pulmonary nodules detection [4] in many CAD systems. Variable N-Quoit filter is based on mathematical morphological technique and formulated as

$$q(x,y) = f(x,y) - f(x,y) \oplus R(x_1,y_1) \tag{1}$$

$$\begin{aligned} &f(x,y) \oplus R(x_1,y_1) \\ &= \max_{(x_1,y_1)\in K_R} \{ f(x-x_1, y-y_1) + R(x_1,y_1) \} \end{aligned} \tag{2}$$

where $(x,y)$ is the coordinate. $f$ and $q$ are pixel values of input and output images, respectively. $R(x_1,y_1)$ is a ring filter function at the target point $(x,y)$. Then the ring filter $R$ is defined by

$$R(x_1,y_1) = \begin{cases} 0, & r_1(x,y) < \sqrt{(x-x_1)^2 + (y-y_1)^2} \le r_2(x,y) \\ -\infty, & \text{otherwise} \end{cases} \tag{3}$$

The ring area $K_R$ is defined by a set of coordinates $(x,y)$ satisfying the condition $R(x_1,y_1) = 0$. $r_1$ and $r_2$ are internal ring radius and external ring radius, respectively. Here $r_2$ is set $r_2 = r_1 + 1$. Note that both $r_1$ and $r_2$ are as functions of the coordinate. The internal ring radius $r_1$ can be changed depended on the target nodule size, and this is a reason why the filter is called "variable" N-Quoit filter. To realize this adaptation, the following distance transformation may be used.

### 2. Distance transformation

Distance transformation is a process to acquire the minimum distance from background pixels (where any pixel value is 0) to each pixel that forms arbitrary target shape. This transformation is defined as

$$d(m,n) = \min_{k,l}(|m-k| + |n-l|) \tag{4}$$

where $d(m,n)$ is output of distance transformation, $(m,n)$ is the coordinate of the arbitrary shape, and $(k,l)$ is of background pixels. We compute this distance by four neighbors in this paper.

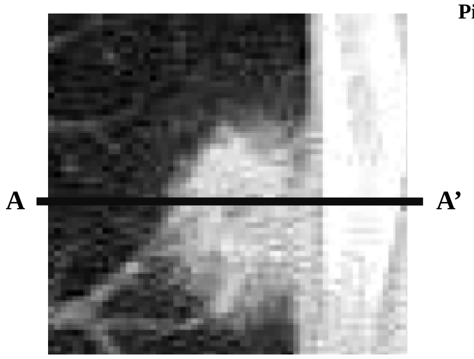
In this transformation, the target area must be distinguished from its background. This distinction can be generally done by a thresholding. It is, however, very difficult to determine the appropriate threshold because it is depended on individual and local properties of each image. In other words, optimal thresholds over a set of images are different from each other. In conventional method, due to this difficulty, detection ability of the filter may be affected badly, especially for non-isolated nodules. Indeed, non-isolated nodules are hardly detected by the N-Quoit filter because of the connecting structures. To overcome this problem, we will propose a new thresholding method in next section.

## III. PROPOSED THRESHOLDING METHOD

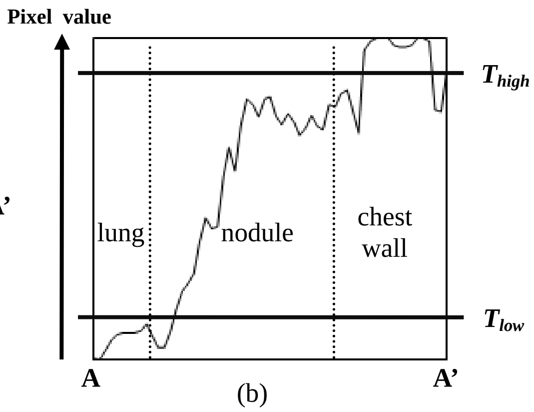
Fig. 1(a) shows an example of a non-isolated nodule that cannot be detected by the conventional N-Quoit filter with the single thresholding technique. The fundamental idea of the proposed thresholding technique is to eliminate the effect of the connecting structures, the chest wall in this case, on the detection ability of the N-Quoit filter. To achieve this, we first consider the distribution property of the pixel value along the cross line A-A' over the undetectable nodule as shown in Fig. 1(a). Fig. 1(b) shows changes of pixel intensity as a function of the coordinates along the cross line A-A' shown in (a). From this figure, we can see that if we use double thresholdings instead of the conventional single one, we can separate the nodule area from both the lung and chest wall area. This process is defined as follows

$$I_{bi}(x,y) = \begin{cases} 1, & T_{low} < I(x,y) \le T_{high} \\ 0, & \text{otherwise} \end{cases} \qquad (5)$$

where $I_{bi}(x,y)$ is output image, $T_{low}$ and $T_{high}$ are lower and higher thresholds, respectively. $I(x,y)$ is the pixel value of intput at the coordinate $(x,y)$. The image acquired by this process is shown in Fig. 2(b), and the input image in Fig. 2(a). Two thresholds were set as following $T_{low}$ = 100 and $T_{high}$ = 200 in this study.

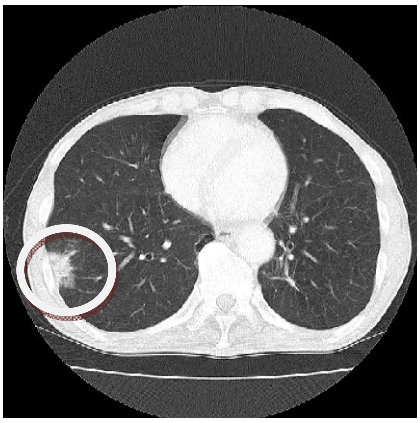



(a)

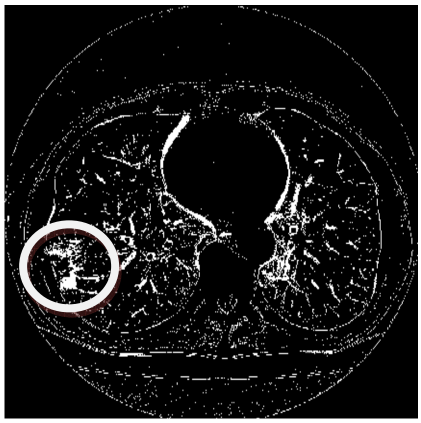



(b)

Fig. 1. (a) Example of a non-isolated nodule. (b) Changes of pixel intensity as a function of the coordinates along the cross line A-A' shown in (a).



(a)



(b)

Fig. 2. (a) Original image has a non-isolated nodule in the white circle. (b) Binary image of the condition $T_{low}$ = 100 and $T_{high}$ = 200.

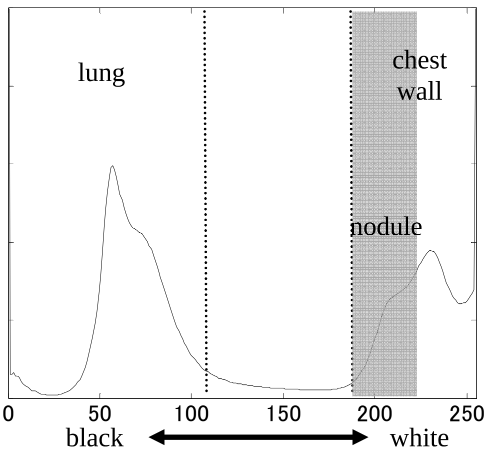



Fig. 3. Average of histogram of 169 chest CT images. Nodules may be included in gray area.

## IV. RESULTS

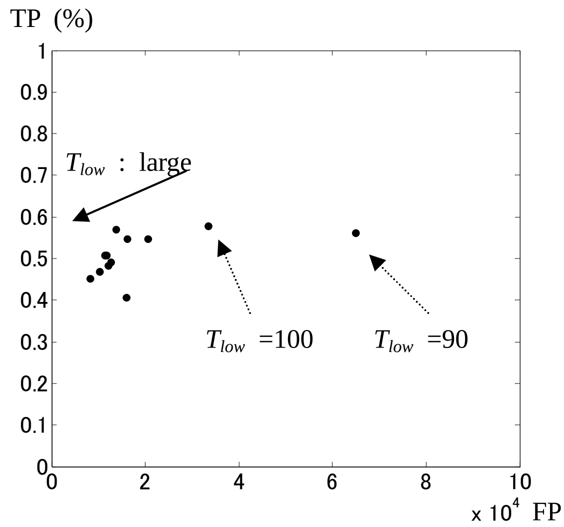
We evaluated the performance of CAD system using receiver operating characteristic (ROC) analysis by changing the parameter of binarizing thresholds

Fig. 3 shows a histogram of 169 chest CT images, and pixel values of nodules are included in the gray area. Here, we need to discriminate nodules from lung and chest wall. We then choose thresholds so that intensities of nodules are greater than the lower threshold $T_{low}$ and less than the higher one $T_{high}$. If $T_{low}$ is chosen around 100, the lung and other structures will be divided. However, it is difficult to choose an appropriate value of $T_{high}$ for discriminating chest wall from other structures as shown in Fig. 3. We thus evaluated the detection performance by using a range of $T_{high}$ that cover the gray area: $T_{high}$ = {190, 200, 210, 220}.

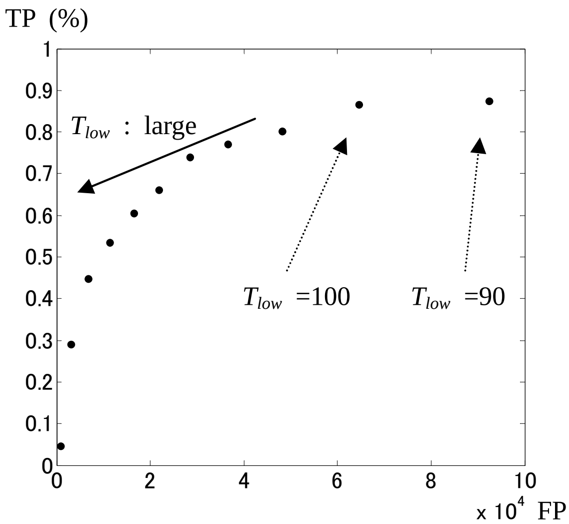
Conventional ROC curves were written as parameter of a threshold (corresponding $T_{low}$ of our method). Then, for comparison, we used a set of parameter values of $T_{low}$ = {90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190}.

Fig. 4(a) shows the ROC curve using the conventional single thresholding technique. While Figs. 4(b) and (c) show the ROC curves using proposed thresholding technique with $T_{high}$ = 200, and 210, respectively. From the results, the best of true positive (TP) is approximately 0.6 in Fig. 4(a) while it is approximately 0.9 in Figs. 4(b) and (c). This result suggests that the proposed technique can be effective for improving the sensitivity of the detection.

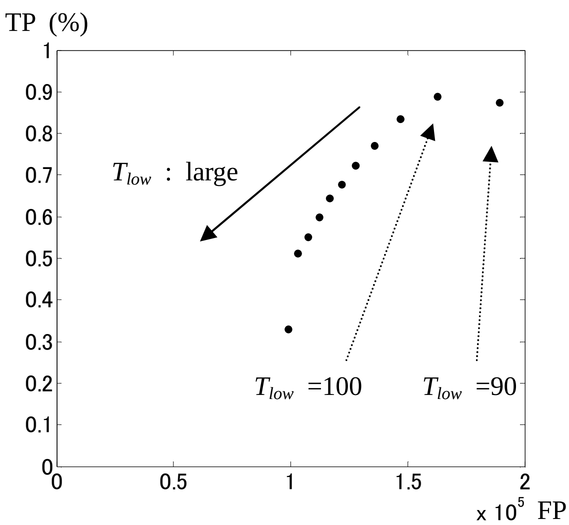
There are, however, a larger number of false positive (FP) in Fig. 4(c) than in Fig. 4(b). This would be caused by isolated areas newly generated in the case that binarizing threshold $T_{high}$ becomes approximately equal to pixel values of the chest wall. In other words, as $T_{high}$ enlarges, the number of pseudo-isolated areas in chest wall increased. Note that the isolated areas in chest wall are not nodules but noises, and thus such areas make FP increase. Fig. 5 as expected from the




(a) Conventional method.




(b) ROC when $T_{high}$ = 200.




(c) ROC when $T_{high}$ = 210.

Fig. 4. ROC curves

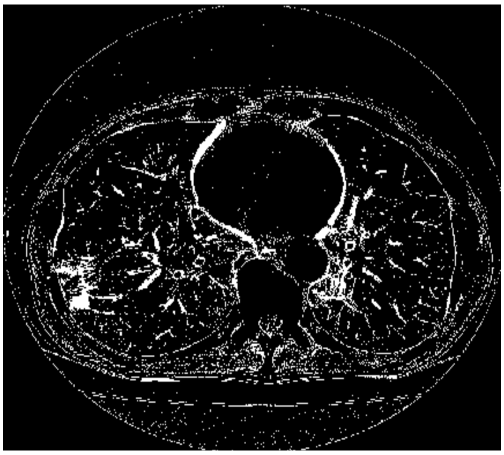


Fig. 5. Many pseudo-isolated areas have generated by an extra large threshold $T_{high}$. ( $T_{low}$ = 100 and $T_{high}$ = 210).

result in Fig. 4(c), shows the binary image in case of $T_{low}$ = 100 and $T_{high}$ = 210. We can see that the number of noise points in Fig. 5 increase compared to Fig. 2(b). we used a smaller threshold than 200, such as $T_{high}$ = 190, TP is getting worse. On the other hand, a larger threshold than 210, such as $T_{high}$ = 220, the detection performance are almost same as for $T_{high}$ = 210. (results area omitted.) From the results, The double thresholding method with $T_{low}$ = 90 or 100 and $T_{high}$ = 200 demonstrates the best performance in our date set.

## V. CONCLUSION

In this paper, we proposed a novel preprocessing technique to improve detection sensitivity in CAD system. The technique is able to divide non-isolated nodules and connecting structures by using two thresholds. We have not conducted to decrease FP yet, but conventional effective methods [5] to do so can be incorporated into the proposed technique.

## VI. ACKNOWLEDGMENTS

This work was partially supported by The Ministry of Education, Culture, Sports, Science and Technology under Grant-in-Aid for Scientific Research #19500413.

## REFERENCES

[1] Takashi Iinuma, Yukio Tateno, Toru Matsumoto, et al. (1992), Preliminary Specification of X-ray CT for Lung Cancer Screening (LSCT) and its Evaluation on Risk-Cost-Effectiveness (in Japanese). Radiation Medicine 52(2): 182-190

[2] Yukio Tateno, Takashi Iinuma, Toru Matsumoto, et al. (1990), Development of the X-ray CT for lung cancer examination (in Japanese). New Medicine in Japan 17 (10):28-32

[3] Shinji Yamamoto, Ippei Tanaka and Masahiro Senda (1993), Image Processing for Computer Aided Diagnosis in the Lung Cancer Screening System by CT (LSCT) (in Japanese). IEICE TRANSACTIONS on Information and Systems J76-D2 (2): 250-260

[4] Tomoko Miwa, Junichi Kako, Shinji Yamamoto, et al. (1999), Automatic Detection of Lung Cancers in Chest CT images by the variable N-Quoit Filter. The transactions of the Institute of Electronics, Information and Communication Engineers J82-D-Ⅱ(2):178-187

[5] Noriyasu Homma, Kazunori Takei, and Tadashi Ishibashi (2008), Combinatorial Effect of Various Features Extraction on Computer Aided Detection of Pulmonary Nodules in X-ray CT Images. WSEAS Trans. Information Science and Applications 5(7):1127-1136

[6] Kazunori Takei, Noriyasu Homma, Tadashi Ishibashi, et al. (2008), Computer aided diagnosis for pulmonary nodules by extracting new shape features from X-ray CT images (in Japan). Japan Society for Fuzzy theory and Intelligent Informatics 20(1):108-116