A Proposal of Pseudo Normal Image Subtraction for Lung Nodule Detection

Masayuki Moribe^{a,*}, Yuriko Tsunoda^a, Hideaki Orii^b

Hideaki Kawano^b, Hiroshi Maeda^b

^a Department of Electrical Engineering and Electronics, Graduate School of Engineering,
Kyushu Institute of Technology, Fukuoka, Japan

^b Department of Electrical Engineering and Electronics, Faculty of Engineering,
Kyushu Institute of Technology, Fukuoka, Japan

*Corresponding Author: m349430m@tobata.isc.kyutech.ac.jp

Abstract

In this paper, a novel computer aided detection (CAD) system for lung nodules in a simple chest X-ray is discussed. In the presence of test subject's past image, the lung nodules can be detected comparatively easily by using temporal subtraction method. However, the temporal subtraction cannot be applied in the absence of test subject's past image. In this study, a normal, i.e. focus-free, image at the time of inspection is estimated from a database composed of different subjects' chest X-rays that have been already diagnosed as normal by medical specialists. We call the estimated image as pseudo normal image. We propose a CAD system based on subtraction between the original test image and the pseudo normal image, and also call the CAD system as pseudo normal image subtraction. As a result of experiments with the use of available simple chest X-rays, the relative advantages against for an existing method were suggested. And it became apparent that the proposed method is a promising approach.

Keywords: lung nodule detection, computer-aided detection, CAD, image synthesis.

1. Introduction

In 2011, in Japan, cancer was accounted for about 28.5% of all deaths. And it was the leading cause of death. Among them, lung cancer was accounted for about 19.7% of all cancer deaths. It was the No.1 cause of cancer death in male and No.2 in female. To reduce the number of lung cancer deaths, it is important to find early-stage lung cancer and cure properly. If operation is delayed and the cancer spreads to other parts of body, it will become difficult to

Fig.1 Simple chest X-ray image and abnormal shade

remove all nidus. If patients suffer from emphysema or bronchitis, it is difficult to remove lungs by surgery. However, it is very difficult to find early-stage lung cancer from subjective symptoms because there are no symptoms such as a cough and a pain.

In order to find early-stage lung cancer, doctors make diagnoses with a visual check using a chest X-ray (Fig.1) or chest X-ray CT image. However, finding the shadows of lung cancer is difficult because the lung nodules to be detected are hidden by bones or normal internal structures such as a heart, a stomach, and so on. In addition, the result of visual inspection is different depending on the doctor's experience or skill. Moreover, heavy burdens are placed on the doctors because they have to check a lot of X-ray images. In recent years, it is often reported the overlooked lung cancer at past X-ray inspection has found in later severe stage. Therefore, the computer-aided detection (CAD) supporting to find lung cancer is expected.

2. Related works

Lung nodules in a chest X-ray image are blurred around the edges and hard to be recognized, thus methods to emphasize the nodules are studied actively. Existing methods include a method with some image filters and a method with bone eliminations. And another existing approach includes subtraction-based methods: the temporal subtraction, the contralateral subtraction, and the similar image subtraction.

2.1 Image Filters

Image filters emphasize lung nodules in a chest X-ray. The image filters include the directional contrast filter for nodule[1], the average radial gradient filter[2], and so on. In general, the image filters is used as a preprocessing of discriminant analysis, because extracted regions with the image filters often contain a lot of different stuffs than lung nodules. Learning-type classifiers or discriminators are often used for the discriminant analysis, e.g. artificial neural networks (ANN). To classify true positive and false positive, for example, we feed the area of an extracted region, circularity, and irregularity into the ANN. However, it is hard to identify lung cancer nodules because the lung nodules present various configurations. As a result, false positives are most likely to be detected in these approaches.

2.2 Energy Subtraction

This method requires two X-ray pictures for the same target under the different energy distribution, and divides two different tissues the individual image. This is caused by the fact that X-ray absorption characteristic varies according to substances. This phenomenon makes it possible to reveal or erase a certain specific substance. By taking advantage of this characteristic, the picture including only calcification shadows, such as a bone tissue (bone image), and the picture including only soft tissues can be obtained. Since the soft tissue image is eliminated bone tissues from a whole image, the soft tissue image is used in diagnosis of a solitary tumor shadow and lung cancer nodules that overlaps the bones. However, because the energy subtraction device is very expensive so far, the penetration rate in ordinary hospitals is quite low. And this device also impose large burden on examinees by increasing the number of radiation exposure. Furthermore, indeed the device can eliminate shadows of bones, but it is still hard to detect lung cancer nodules hidden by shadows

of blood vessels.

2.3 Contralateral Subtraction

The Contralateral subtraction proposed by Li et al. [3][4] is a method to emphasize lung nodules by using the fact that the lung of human is approximately symmetric. At first, the rib cage boundary is detected in a chest X-ray image. Next, the detected lung region determined by the rib cage boundary is translated and rotated to match the midline derived from the rib cage with the perpendicular center of image. Then, a mirror-reversed image is generated according to the perpendicular center. Finally, it is possible to emphasize the part of abnormal shadow by inspecting the subtraction between the original test image and the mirror-reversed image as shown in Fig.2. However, because the position of normal structures such as organs differs slightly by left and right in actual, the subtraction image includes a lot of normal structures.

2.4 Similar Image Subtraction

The similar image subtraction is a method which uses similar other person's chest X-ray image instead of the past image of the same examinee to detect lung cancer nodules by using only one current image as same as the contralateral subtraction. In the similar image subtraction, the database that is composed of normal chest X-ray images obtained from a lot of people is prepared beforehand. First, a similar image to the test image is found automatically among the database. Next, the similar image is non-linearly deformed to match with the test image more. Then, the subtraction image between the test image and the deformed image is generated. Oda et al.[5] selected 4,000 images according to the age and the sex from the database. The 4000 images were narrowed to 100 images according to the area and height of the lung. Finally, the most similar image was found from the 100 images. However, the similar image which is created by only one database image does not fit at some parts.

3. Proposed Method

In the previous section, we indicate that the existing methods have some demerits and the detection rate is not so high. In this paper, we propose a novel method to develop a pseudo normal image that is similar to the test image by local matching from a lot of images in database. The local matching could bring a quite similar image over the image in detail. Moreover, because the proposed method does not

Fig.2 algorithm

contain any deformation, the structure of similar image is preserved. As a result, the detection rate could be increased.

3.1 Algorithm

At first, database of normal chest X-ray image is prepared. It is already diagnosed by doctor and there were no nodules. The large number of images and wide range of age or sex are desirable. In the experiment, we used 87 images as database image.

3.1.1 Creation of Pseudo Normal Image

Database is searched to decide similar local area for test image. test image patch(p [px] * q [px]) is cut around the origin of images which is upper left. The x-axis is horizontal direction and the y-axis is vertical direction. The similar image patch (p [px] * q [px]) is searched in the search area (s [px] * t [px], $p \leq s$ and $q \leq t$) at the database image. This process is repeated all database images and decide similar image patch in each database image (Fig.2). The most similar image patch is selected from similar image patches because of the highest degree of similarity. The mean of the most similar image patch is matched the mean of the test image patch. The center of test image patch moves to k [px] toward the x-axis and decide the most similar image patch. If the center of test image patch is the right endpoint of the test image, it moves to k [px] toward the y-axis. This process is repeated until the center of test image patch reaches the lower right of test

Fig.3 Multiple threshold

image. The most similar patches for the test image are decided by these processes and the pseudo normal image is created from these patches. The most similar patches have duplicative area because the target patches have it. The area is used the mean of the corresponding pixels. The pseudo normal image which is corresponding the test image is created from this way.

3.1.2 Detection of The Abnormal Shade

Generally, the pixel value of lung nodule is higher than other normal parts. Therefore, lung nodule is emphasized by subtracting the pseudo normal image from the test image. The nodule is detected from the subtracted image by binarization which uses the multiple-thresholding technique(Fig.3). The following is the process of the multiple-thresholding technique. First, the histogram about the pixel value is calculated from the subtraction image. The multiple-thresholding technique uses the some percentages of the histogram area. The percentages are set in advance. Second, the highest of pixel value is start point and the threshold values are decided the pixel values which measure up to setting percentage of the histogram area. Finally, the binarized images are created from the subtracted image used the threshold values. Each of binarized images is labeled as connected region and island regions are detected binarized images. Some island regions have duplicative area which is contained each other. The binarized images are checked whether they have island regions contained each other or not. If the binarized images have contained island regions, leave the only small one. This process, it is possible to leave island regions have higher pixel value (Fig.4). The result image is decided the initial candidates image. The island regions from the initial candidates image is calculated degreed of circularity and removed from initial candidates as false positive when the degreed of circularity of the island region under the threshold of it.

3.2 Template Matching

Template matching is the search method (Fig.5).

Fig.4 Detection of abnormal shade candidates

Fig.5 template matching

Template image $T(N_T * M_T)$ is moved in search area $I(N_I * M_I)$ and decided the position which the degree of similarity is higher than others.

3.3 Degree of Similarity

In this paper, we use the normalized correlation coefficient R to calculate degree of similarity at the template matching. If inner product is calculated between with one's fellow vectors, the correlation coefficient R is $\cos \theta$. The following, the equation of normalized correlation coefficient R is shown.

$$R = \frac{\sum_{i=0}^{M_I-1} \sum_{j=0}^{N_I-1} (I(i, j) - I) * (T(i, j) - T)}{\sqrt{\sum_{i=0}^{M_I-1} \sum_{j=0}^{N_I-1} (I(i, j) - I)^2} \sqrt{\sum_{i=0}^{M_T-1} \sum_{j=0}^{N_T-1} (T(i, j) - T)^2}} \quad (1)$$

$$I = \frac{1}{M_I N_I} \sum_{i=0}^{M_I-1} \sum_{j=0}^{N_I-1} I(i, j), T = \frac{1}{M_T N_T} \sum_{i=0}^{M_T-1} \sum_{j=0}^{N_T-1} T(i, j) \quad (2)$$

Therefore, the range of R is -1 to 1. If the value of R is near the 1, this image has positive correlation, and if the value of R is near the -1, this image has negative correlation. Normalized correlation coefficient is able to stably calculate degree of similarity even if the brightness has ups and downs. Therefore, it can use position adjustment of two images, even if the images have difference of contrast or brightness.

3.4 Degree of Circularity

It is calculated degree of circularity C and valid diameter R to decide lung nodule candidate. They are represented by area of nodule S and area of nodule within circle which with equivalent nodule area A . it is shown by equation (3), (4).

$$C = \frac{A}{S} \quad (3)$$

$$R = 2 \sqrt{\frac{S}{\pi}} \quad (4)$$

4. Experiment

In order to verify the effectiveness of the proposed method, a comparative experiment was conducted.

4.1 Experimental Setup

In the experiments, 87 normal images of Standard Digital Image Database created by Japanese Society of Radiological Technology[6] were used as the database of normal chest X-ray images. In addition, 50 abnormal images were used as the test images. Database images and test images has already been diagnosed by doctors whether a lung cancer is found or not. The number of normal images of database is 93 in actual. However, 6 of 93 images (JPCNN003, 007, 044, 051, 077, and 083) contain unusual objects such as a shadow of medical equipment and sewing works. Therefore, we used 87 images as database. All of the database images are 2048 * 2048 [px], 0.175mm of pixel interval and 4096 gradations (12 bit). In the experiment, we resized database images to 256 * 256 [px] and 256 gradations to reduce a computational time. The parameters to create the pseudo normal image were set as follows: The patch size cutting from test image was set as 45 * 45 [px], and the search area was 51 * 51 [px]. Central coordinate of the test image patch for trimming away was moved every 4 [px]. Next, we got a subtraction between test image and the created pseudo normal image, and the shadow of lung nodule was detected. The subtracted image was created by subtracting gradation of test image to gradation of pseudo normal image. When we showed subtracted images on screen, the constant gray value 128 was added to the each gray value of subtracted images. The gradation of lung cancer is higher than other parts, so the shadow of lung cancer should remain in subtracted image. Taking advantage of this, the subtracted image was binarized. After

binarization, candidates of abnormal shade were detected. In the binarization, the threshold values were set as every 4 percentile value from 4 to 20, and the threshold value of circularity degree was set as 0.45. In the experiment, the detection rate is defined as a ratio of success samples against for the whole test images. When the detected regions cover the true abnormal shade, we assumed the detection as success.

4.2 Experimental Results

As a result of experiments mentioned above, the proposed method could detect 41 test images correctly among 50 test samples. The detection rate is calculated as 82%. As an example of test images, the result of test image JPCLN006 is shown in Fig.6. In Fig.6, the images (a), (b), (c) shows the test image itself including the mark of cancer position, the created pseudo normal image in the proposed method, the subtraction image between the test image and the pseudo normal image, respectively. As we can see, the abnormal region presents relatively high gray value compared to the surrounding area as an evidence of abnormality. Meanwhile, 50 of 73 test images could be detected by using the existing method according to the Oda's paper[5]. In this case, the detection rate of the existing method can be calculated as 68%. While the experimental settings between our experiments and Oda's experiments, we believe the superior position in comparison of detection rate suggests the effectiveness of the proposed method.

5. Conclusions

In this paper, we proposed the CAD system using one simple chest X-ray image to support doctor's diagnosis. In the experiment, the pseudo normal image was generated by using our proposed method. In addition, abnormal shade is detected in test image. As a result of experiment, this method achieves a higher detection rate than existing method. However, the pseudo normal image made by this method is indistinct than the test image. Therefore, there are some cases that the subtracted image has many noises. In addition, the number of database image is very low at 87. Some existing method uses 14,564 for database image. To increase database images is necessary because prospect which contains more similar patch is increased. From now on, we increase the number of database and improve the degree of similarity and detection method to make higher

Fig.6 result images.

precision pseudo normal image.

References

- (1) H.Suzuki, N.Inaoka, H.Takabatake, M.Mori, S.Sasaoka, H.Natori and A.Suzuki : "A Computer- Aided System for Lung Nodule Detection on Chest X-ray Images", Medical Image Technology, Vol. 10, pp. 17-22, 1992.
- (2) J.Shiraishi, Q.Li, K.Suzuki, R.Engelmann and K.Doï : "Computer aided diagnostic scheme for the detection of lung nodules on chest radiographs : Localized search method based on anatomical classification", Medical Physics, Vol. 33, No. 7, pp. 2642-2653, 2006.
- (3) Q.Li, S.Katsuragawa and K.Doï : "Imoroved contralateral subtraction images by use of elastic matching technique", Medical Physics, Vol. 27, No.8, pp.1934-1942, 2000.
- (4) Y.Harada, S.Kido, H.Shouno and S.Kakeda : "A Contralateral Subtraction Scheme for Detection of Pulmonary Nodules in Chest Radiographs", IEICE Technical Report, MI2009-55 , 2009.
- (5) N.Oda, T.Aoki, H.Okazaki, S.Kakeda, Y.Kourogi, K.Yahara and H.Shouno : "Development of Computerized System for Selection of Similar Images from Different Patients for Image Subtraction of Chest Radiographs", JSMBE, Vol.44, No.3, pp.435-444, 2006.
- (6) Japanese Society of Radiological Technology : "Standard Digital Image Database:Chest Lung Nodules and Non-nodules", 1998.