Automated Lesion Detection Methods for 2D and 3D Chest X-Ray Images

Takeshi Hara, Hiroshi Fujita,Yongbum Lee, Hitoshi Yoshimura* and Shoji Kido**
Department of Information Science, Gifu University
Yanagido 1-1, Gifu, Gifu 501-1193, JAPAN. TEL: +81-58-293-2757, FAX: +81-58-230-1895
{hara, fujita, lee}@fjt.info.gifu-u.ac.jp
*Center Research Laboratory, Konica Co.
**Osaka Medical Center for Cancer and Cardiovascular Diseases

Abstract

The purpose of this work is to present some technical approaches of our computer-aided detection (CAD) system for chest radiograms and helical CT scans, and also evaluate that by using three databases. The CAD includes some methods to detect lesions and to eliminate false-positive findings. The detective methods consist of template matching and artificial neural network approaches. Genetic Algorithm (GA) was employed in template matching to select a matched image from various reference patterns. Artificial Neural Networks (ANN) were also applied to eliminate the false-positive candidates. The sensitivity and the number of false-positives were 73% and 11FPs per image on chest radiogram CAD and 77% with 2.6 FPs per image on helical CT scan CAD. These preliminary results imply that the GA and ANN based detective methods may be effective to indicate lesions on chest radiograms and helical CT scans.

1. Introduction

Early detection of suspicious lesions is effective to reduce the number of death caused by lung cancer. Although a conventional chest radiogram has been used for the screening of lung cancer, mass screening by helical CT scan has started for the high risk groups of smokers in some facilities of Japan. It shows a good performance for the diagnosis. We are currently working on developing two CAD schemes for chest radiograms and helical CT scans. The purpose of this study is to show recent results for chest radiogram CAD system by using three different databases and to provide a preliminary detection method for CAD system on chest helical CT scans.

2. Materials and methods

2.1 CAD system for chest radiograms

Conventional chest radiogram has been employed as a mass screening method of lung cancer for decades, however it was said that small lesions at the first stage of cancer could appear with a very subtle shadows that radiologists might be miss. We have been developing a CAD scheme by computer programs that could prepare quantitative values and the position of lesions to radiologists. If radiologists will take into account the information from CAD system, their diagnostic performance would be higher when the information was not prepared. Figure 1 shows the overview of our

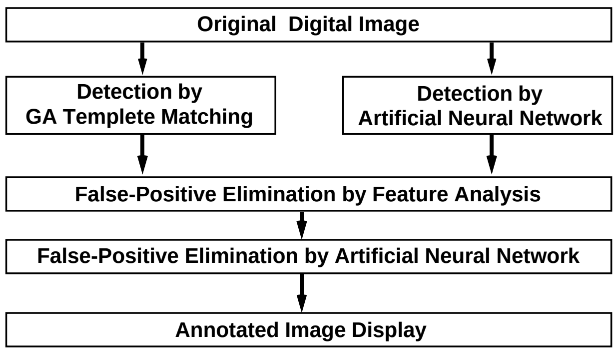


Fig.1 Overview of chest radiogram CAD

The detection scheme consists of two different part: detection and false-positive elimination processing. Many lesion candidates were indicated by two detection process by GA template matching and ANN technique not to miss true-positive findings, however some part of those candidates include normal shadows as false-positive findings.

automated detection system. Each process of this scheme was explained as follows.

2.1.1 Image database

We employed three image databases to estimate our CAD scheme. Table 1 shows the number of images in each database. Images in Group-A are collected in a prefectural mass survey in Japan and include relatively easy cases with large nodules. Group-B is a standard database published by Japanese Society of Radiological Technology (JSRT) [1]. The quality is excellent to evaluate the CAD performance because the ease of detection for each case was already classified into five categories and verified by ROC studies. Group-C is mass survey cases and is collected in a central hospital in Osaka, Japan. All the films were digitized at 0.1mm or 0.175mm with 10-bit or 12-bit density resolution. We converted all of the digitized images to 1024 x 1024 pixels with 10-bit density resolution.

2.1.2 Lung field extraction

The lung field was extracted to limit the region to be processed. Thresholding techniques for gray scale images were applied to extract the rough region at first. The threshold value was determined in terms of the area of histogram of original images[2]. Larger regions that correspond to the lung field were almost extracted in this step. Secondly, smaller regions that could not be considered as lung fields were eliminated in terms of the region size. Finally, the extracted lung regions were expanded because the extracted area could not always include borders of lung field. Figure 2 shows an example of an original chest image (a) and the processed image (b). The lung area with margins was extracted correctly.

2.1.3 Nodule detection by GA and TM

Genetic algorithms (GA) is a optimal value search procedure in complex problems[3, 4]. We apply the GA technique to the detection process as a pattern recognition methods based on template matching with various reference patterns[5]. Detection method at this step is based on template matching techniques with many reference images. We employed a two-dimensional Gaussian distribution for some reference ones because the shadow of lung nodules in radiogram generally shows the distributions. The size or angle was unknown in practice, however, we prepared various patters of Gaussian distribution.

Table 1 Employed three databases

	# normal	# abnormal	total
Group-A	10	33	43
Group-B	93	84	177
Group-C	59	28	87
Total	162	195	357

(a) (b)

Fig.2 Lung field extraction

Example of original chest image (a) and extracted image (b). The precious lung area was enlarged not to miss the nodules along the lung wall.

Here, a GA technique was employed to select the suitable image from various reference patterns and to show the fittest position in the original chest image. Virtual creatures, which is called individuals, have chromosomes represented by a binary digit system and the chromosomes could determine the pattern and the position by calculating the ease of living of individuals in the environment. The degree of living was given by a fitness defined as a difference between a reference image (Gaussian image) and an observed image (extracted patients' image). The genetic operations including crossover and mutation were applied to all individuals to create new ones in the next generation.

Figure 3 shows the overview of detection step based on GA template matching method. We prepared 512 reference images with various sizes and angles of Gaussian distributions as shown in the left half. The chromosomes determine the positions x and y of partial image as shown in the right half. The similarity was defined as a five-dimensional function with those parameters. The chromosomes have 42 bits that represent the recognized positions of x (10 bits: 0 to 1023) and y

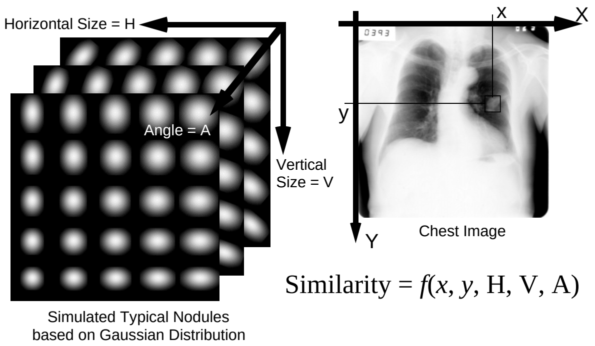


Fig. 3 Detection by using GA template matching

The position of nodule candidates on a chest image is defined by two parameters of x and y (Right). Three parameters of H, V and A determines the shapes of simulated typical nodules (Left). A similarity is calculated in terms of those five parameters, so that the recognition of nodule existence is an optimal value search problem in those five dimensional space.

(10 bits: 0 to 1023), horizontal and vertical size categories of H (8 bits: #1 ~ #8) and V (8 bits: #1 ~ #8) and the category of angle of reference images A (8 bits: #1 ~ #8).

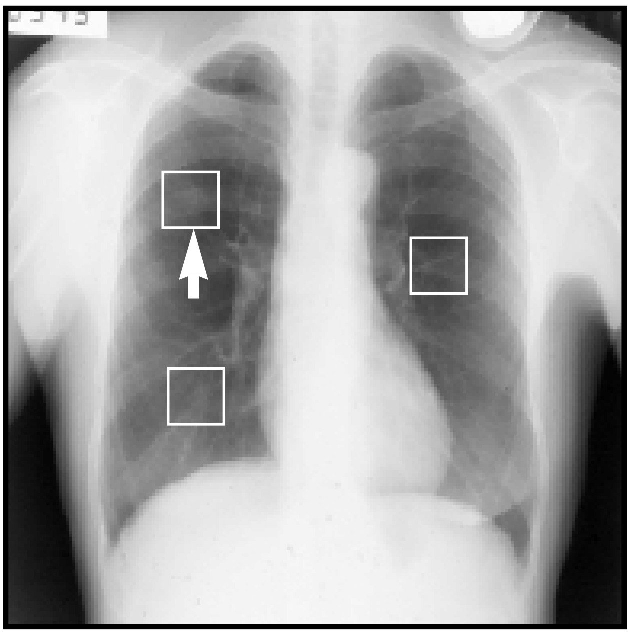
The population size, generation, and probability of mutation were 120, 200 and 0.25 to 0.01 as the generation gains. The mutation was set up by inverting all bits at the probability. Then, ranking selection and one-point crossover technique were employed in the GA procedures.

2.1.4 Nodule detection by ANN

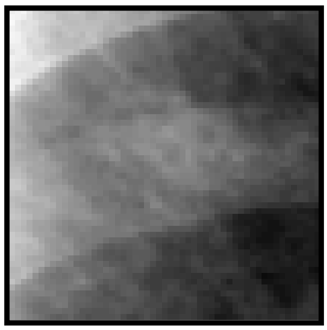
ANN based detection method includes feed-forward artificial neural network with three layers. The regulated pixel values in terms of the minimum and maximum in a partial image were entered to ANN. The matrix size of inputted image was 32 x 32 and the gray scale values were converted to 0.0 to 1.0. To train the ANN by back-propagation method, we selected a hundred diagnosed shadows of fifty of nodular and fifty of obvious normal. The iteration and the errors for training were 2000 and 0.0001, respectively. By scanning the mask of 32 x 32 with eight steps, the output value of ANN was recorded, and the positions of candidates that satisfy certain conditions of output values were remained as the detection region for nodular shadows.

2.1.5 Elimination of false-positive candidates by feature analysis

Candidates that were over-interpreted are called false-positive candidates. We previously analyzed the image and found that the false positive candidates could be classified into three categories such as shadows of lung margins, shadows of bones and shadows of vessels and others. We developed the algorithm to eliminate those candidates selectively. The first classification step was based on the histogram analysis within the candidates. The candidate that includes many regions outside expanded lung field was eliminated in terms of the area of low pixel values in its histogram. The second classification step was based on the histogram analysis of



(a)



(b)

Fig.4 Detection example

Computer and radiologist's detection were superimposed with three white squares and a white arrow on lung image (a). Upper left square (Upper right lung) was detected correctly, but the others were miss detection (false-positives) by computer. (b) is the magnified image of radiologist detection (ROI indicated by white arrow). The nodule size is medium and it is said that the shadow is not so subtle to diagnose.

the vector images within the candidates. The kurtosis that is the feature value of the sharpness of histogram was extracted. The final classification step was based on the analysis of feature values of the entropy and the ratio of partial images inside the candidates.

2.1.6 Elimination of false-positive candidates by ANN

This process was very similar to the ANN-based-detection step. The remaining candidates were entered to an another ANN. To train the ANN to eliminate false-positive candidates, we selected a hundred diagnosed shadows of fifty of nodular and fifty of normal. The conditions of iteration and the errors were the same as that of the detection step. The number of nodes of input layer, hidden layer and output layer were 32 x 32, 32 x 32 and 1, respectively. Normalize pixel values within partial images were calculated and entered to ANN.

2.1.7 Indication of detected candidates

Detected candidates were indicated by squares on the original images with the size of 32x32. Figure 4 shows two sample images with nodules indicating by white arrows. There were two miss detection, called false-positives, but a small nodule shadows was detected correctly.

2.2 CAD system on chest helical CT scans

A similar procedure could be applied to the detection part of the nodules on chest helical CT scan images, and the diagnosis by CT images play an important role in not only confirming the existence of nodules but also discriminating. We have been developing an automated detection scheme to detect nodular shadows on CT scans. Figure 5 shows the overview of our automated detection system. Each process of this scheme is explained as follows.

2.2.1 Image database

Table 2 shows the specification of chest helical CT scans used in this study. The datasets consist of 20 cases (5 normal and 15 abnormal cases) including 101 nodules. The abnormal cases have multiple nodules, so that a few nodules with various size could be seen in the same slice images.

2.2.2 Detection methods by GA and TM

The detection technique is based on GA and template matching method (TM) in a three-dimensional space. We supposed that it was possible to detect a nodule to set the higher adaptation scales surrounding the nodules, so that the individuals could gather on the place for the environmental configuration. We used a similarity, which was defined by the difference of CT values between three-dimensional images, for adaptation scales of individuals. The similarity was calculated by template matching between two images.

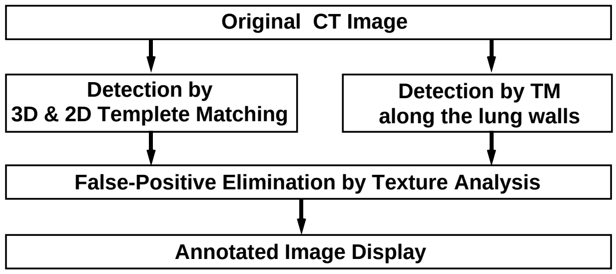


Fig.5 Overview of chest CT CAD

The detection scheme consists of two different part: detection and false-positive elimination processing. Many lesion candidates were indicated by two detection process by GA template matching and template matching along the lung walls not to miss true-positive findings, ant the false candidates were also detected as chest CAD scheme, so that the false-positive elimination process was attached after the detection process.

Template matching is one of the most fundamental

Table 2 Specification of data set

Tube voltage	120 [kV]
Tube current	50 [mA]
Slice thickness	10 [mm]
Scan time	1 [sec/round]
Table feed rate	20 [mm/sec]
# of cases	20
# of images	557
# of nodules	101

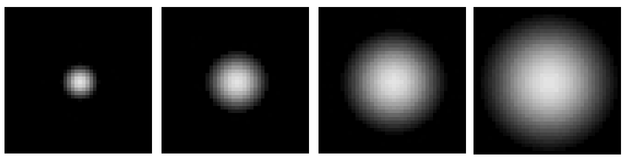


Fig.6 Employed four reference images

The sphere size is 10, 20,30 and 40 pixels (left to right). The information of chromosomes determines automatically to employ which reference pattern in four based on the fitness values.

methods to detect a target object within an image field. If the similarity between an unknown object (observed image) and the template (reference image) is sufficiently high, the unknown object can be considered as the detection object. We used three-dimensional images of the chest helical CT scans and simulated nodule images for observed and reference images, respectively. Each of the CT scans (observed image) has 512x512 pixels and consists of about 30 slice images.

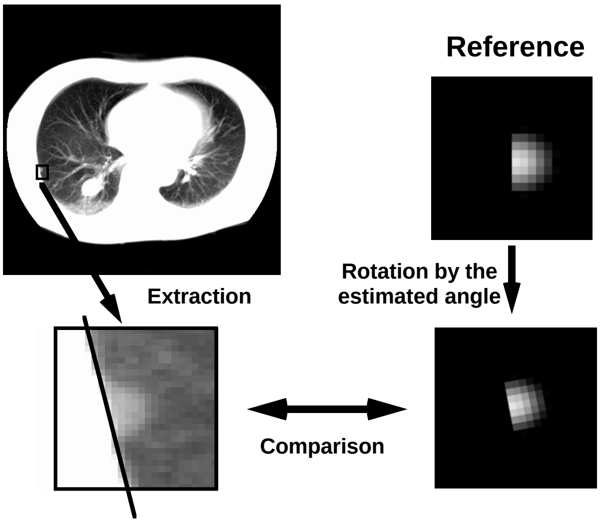
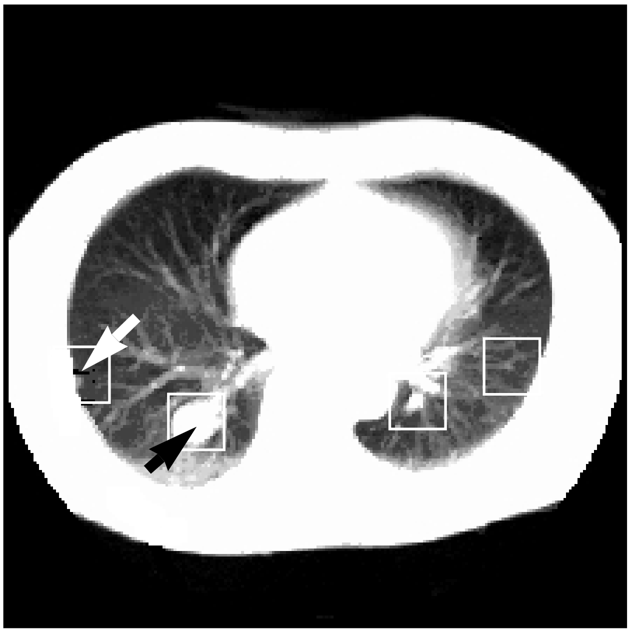


Fig.7 Overview of template matching along lung walls

A reference image with semi-sphere shape was used rotating by the estimated angle of chest walls. The angle was determined by curve fitting techniques on CT slice images.



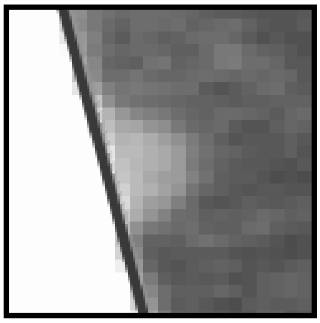
(a)

We made four artificial simulated nodule images which include a similar image to be detected. Here, lung nodule shadow generally shows a spherical shape and the CT values tend to have a Gaussian distribution, we used some simple sphere models with three-dimensional Gaussian distribution as simulated nodules of reference images. Four reference images ranged 3 slices and their diameters were 10, 20, 30 and 40 pixels as shown in Fig.6. The observed image was cut off at a position which was determined by place information in each chromosome of GA [6]. At the same time, each chromosome selected a reference image from four ones.

The size of cut image was the same as the reference image. The similarity between a cut and a reference images was calculated to evaluate adaptation scale of the individual on GA. We used cross-correlation as the similarity. The similarity was only evaluated within the common area that contained only a part of nodules between a cut and a reference images. Therefore it was possible to evaluate the similarity without depending on the background patterns.

2.2.3 Detection method along the lung wall

Conventional template matching technique was also used to detect nodules existing along the lung walls. The GA template matching method was not able to detect these nodules because they usually have semicircular



(b)

Fig.8 Detection example

Computer and radiologist's detection were superimposed with four white squares and black and white arrow on lung image (a), respectively. The white arrow indicates a small nodule on the wall. The radius was about 4mm. The magnified image of the nodule is (b). The black arrow indicates a large nodule. Computer also detected the two nodules correctly, but two false-positive candidates was detected. The black line on (b) is an approximated line from the wall shape.

Table 3 System performance

	TP	# of FP /img
Group-A	78 % (35/45)	6.0
Group-B	68 % (57/84)	16.5
Group-C	82 % (23/28)	11.7
Average	73% (115/157)	11.4

shapes as shown in Fig.7. A reference pattern with semicircular shape was employed for this template matching. The reference pattern was rotated an angle to the lung wall automatically.

2.2.4 Elimination of false-positive candidates by feature analysis

Detected candidates were classified into true or false-positives by texture analysis. Four feature values such as contrast in gray level differences and entropy, inverse difference moment and angular second moment in co-occurrence matrix were employed to eliminate the false-positive candidates.

2.2.5 Indication of detected candidates

Detected candidates were indicated by white square on the original images. Figure 8 shows a detection example with two nodules, one is a large nodule indicated by black arrow and the other is a very small nodules on the lung wall. The radius of it was approximately 4mm because the actual pixel size was about 0.6mm. The black line on Fig8 (b) was approximated lines from the lung wall shape.

3. Results

3.1 Chest radiogram CAD system

We employed three image databases to estimate the detection performance. Each performance is shown in Table 3. The average of sensitivity and the number of false-positives (FPs) were 73% and about 11 per image, respectively, which require the reduction of FPs for practical use of CAD.

3.2 CAD system on chest helical CT scan

The methods were applied to 20 cases (5 normal and 15 abnormal) including 101 nodules with many smaller sizes less than 10mm. It was possible to detect 78 nodules by combining the GA method and the template matching along the lung walls. However, the number of false-positives was 2.6 per slice.

4. Conclusion

Image recognition method by using Genetic Algorithm is useful for detecting and locating lung nodules on digital chest radiogram and helical CT scans. Three datasets (357 images) were tested to estimate the system performance for chest radiogram CAD system including JSRT standard image database issued in Japan. The detection performance was 73 % true-positive rate with 11.4 false-positives per image.

As for the CAD system on chest helical CT scans, our detection methods identified 77% of true-positive nodules with 2.6 false-positives per image, as a preliminary study with a small dataset.

Acknowledgment

The authors are grateful to J. Shiraishi for his proposal for employing the JSRT database, and to A. Kojima for his helpful discussions and technical assistance.

References:

1. Standard digital image database for chest radiographs with and without a lung nodule (Jun-ichi Shiraishi, Kunio Doi, Shigehiko Katsuragawa, et al.), Proc. of the 12th International Symposium and Exhibition on Computer Assisted Radiology and Surgery (CAR' 98), p.882, (1998)
2. Machine vision: theory, algorithms, practicalities (E.R. Davies) Academic press, (1997)
3. Genetic algorithms + data structures = Evolution programs (Zbigniew Michalewicz) Springer-Verlag, (1992)
4. Advances in genetic programming (edited by Kenneth E. Kinnerar, Jr.) the MIT Press, (1994)
5. Development of automated detection system for lung nodules in chest radiograms (Takeshi Hara, Hiroshi Fujita and Jing Xu) Proc. of IASTED International Conference on Intelligent Information Systems (IEEE Computer Society), ed. by H.Adeli, p.71-p.74 (1997)
6. Nodule detection on chest helical CT scans by using genetic algorithm (Yongbum Lee, Takeshi Hara, Hiroshi Fujita, et al.) Proc. of IASTED International Conference on Intelligent Information Systems (IEEE Computer Society), ed. by H.Adeli, p.67-p.70 (1997)